Multisensory Processing and Perceptual Consciousness: Part I

Robert Eamon Briscoe*
Ohio University

Abstract

Multisensory processing encompasses all of the various ways in which the presence of information in one sensory modality can adaptively influence the processing of information in a different modality. In Part I of this survey article, I begin by presenting a cartography of some of the more extensively investigated forms of multisensory processing, with a special focus on two distinct types of multisensory integration. I briefly discuss the conditions under which these different forms of multisensory processing occur as well as their important perceptual consequences and interrelations. In Part II, I then turn to examining of some of the different possible ways in which the structure of conscious perceptual experience might also be characterized as multisensory. In addition, I discuss the significance of research on multisensory processing and multisensory consciousness for philosophical attempts to individuate the senses.

1. Introduction

Multisensory processing, as understood here, encompasses all of the various ways in which the presence of information in one sensory modality can adaptively influence the way information is processed in a different modality.[1] Multisensory processes are adaptive in the sense that they function to enhance the accuracy of perception and the control of perceptually guided actions.[2] Some forms of multisensory processing, for example, deal with the challenge of tracking or identifying objects and events crossmodally. When sensory signals in different modalities have been identified as having a common origin in the external environment (or the perceiver's body), other forms of multisensory processing address the problem of integrating them and, importantly, reconciling them when they conflict. Some of these operate on very short time-scales, while others involve the slow recalibration of one modality to another.

This will be a two–part survey article. In this first part, I begin by presenting a cartography of some of the more extensively investigated forms of multisensory processing, with a special focus on two distinct types of multisensory integration. Research on multisensory processing, it will become clear, provides reasons to be skeptical that an account of perception as a whole can be constructed on a sense-by-sense basis, treating different perceptual modalities as explanatorily independent building blocks.

In the second part of this survey, I turn to examining of some of the different possible ways in which the structure of conscious perceptual experience might also be characterized as multisensory. To anticipate, perceptual consciousness might be multisensory only in the minimal sense that experiences in different modalities are sometimes simultaneously *co-conscious*. Alternatively, perceptual consciousness might be multisensory in the more robust sense that certain aspects of your overall experience of the world at a given moment in time are not always *specific* to a particular modality. For example, when you see and hear a kettle at boil, the kettle may be represented in a non–modality-specific way as the common object of both vision and audition (O'Callaghan 2012, 2014; Bayne 2014). These are just two of the possibilities to be considered.

Another aim of the second part of this survey will be to examine the significance of research on multisensory processing and multisensory consciousness for philosophical attempts to individuate the senses, that is, to explain what distinguishes one sense from another. As we shall see, the multisensory character of perception significantly complicates attempts to individuate the senses by means of certain commonly applied criteria.

2. Some Varieties of Multisensory Processing

The purpose of this section is to present a cartography of some of the more extensively studied forms of multisensory processing. I briefly discuss the conditions under which these different forms of multisensory processing occur as well as their important perceptual consequences and interrelations.

2.1. MULTISENSORY CAUSAL INFERENCE

Proximal sensory signals underdetermine their distal sources in the world. A given pattern of retinal stimulation, for example, could be caused by the light reflected from many different 3D scenes (this is the so-called 'inverse optics' problem for vision). The fundamental challenge faced by any perceptual system, according to recently influential Bayesian models of perception, is to infer the most probable cause of a given pattern of proximal sensory stimulation on the basis of (1) the low-level cues present in the pattern itself, i.e., properties of the pattern that are predictive of properties in the environment, and (2) pre-wired or learned assumptions about the environment's statistical structure (the perceptual system's 'prior knowledge' about the world). Call this challenge the *unisensory causal inference* problem.[3]

Multiple perceptual systems are very often activated simultaneously, however. Everyday motor actions, like knocking on a door or preparing a cup of tea, typically generate visual, tactile, auditory, and proprioceptive feedback at the same time. In addition to the problem of unisensory causal inference, there is thus also the problem of *multisensory causal inference*, the challenge of determining when sensory signals received in different modalities have the same distal cause and when they do not (Bedford 2001, Körding et al. 2007, Shams & Beierholm 2010, Landy et al. 2011, Shams 2012). Bedford (2001) refers to this challenge as the 'object identity decision', while van Dam et al. (2014) refer to it as the 'correspondence problem'.

Multisensory causal inference is frequently discussed in connection with the *binding problem* in perceptual psychology and neuroscience (Treisman 1999, 2003; Robertson 2003; for philosophical discussion, see Deroy 2014, O'Callaghan 2014, and de Vignemont 2014a). In the unisensory visual case, for example, the problem is that of explaining how features represented in anatomically different visual processing areas of the brain are properly attributed or 'bound' to each of the various objects visible in a scene. In the multisensory case, the problem is that of characterizing those processes in virtue of which features represented by different modalities are attributed to either to the same or different objects. Shams & Beierholm (2010) write: 'At any given moment, an individual typically receives multiple sensory signals from the different modalities and needs to determine which of these signals originated from the same object and should be combined and which originated from different objects and should not be bound together' (425). Multisensory causal inference, they suggest, is the process whereby the multisensory binding problem is solved.

Solving the multisensory causal inference problem involves different types of information. One source of information is the prior (subjective) probability that a set of signals arising in different modalities have a common cause. On the Bayesian approach, this is sometimes referred to as a 'coupling prior' (Ernst 2006, Helbig & Ernst 2007). An additional 'bottom–up' kind of evidence for such inference is *spatio-temporal congruence*. Signals that originate in the same region

of space and at the same time, other things being equal, are more likely to have a common cause than those that don't. One way to characterize spatio-temporal congruence is in terms of the number of 'amodal' attributes shared by objects represented in two or more modalities (Welch 1999). These include: (1) location; (2) shape; (3) size; (4) orientation; (5) texture; (6) kinetic properties; (7) temporal onset; and (8) temporal frequency. These attributes are amodal – or better *multi-modal* – in the sense that they can be perceived through more than one modality. They are, as philosophers would say, 'common sensibles' as opposed to 'proper sensibles'.

2.2. Optimizing Multisensory Integration

Capacities for multisensory causal inference, however they are implemented, are necessary for what I shall here refer to as *optimizing multisensory integration* (or ***O-integration*** for short). The initial estimates of a property produced by different modalities can conflict. For example, audition and vision might produce different initial estimates of the time or location at which an event in the environment occurs. Alternatively, vision and touch might produce conflicting initial estimates of an object's shape or orientation. Such intersensory discrepancies can arise for a variety reasons. One modality may simply have a finer grain when it comes to discriminating the relevant property. Initial estimates may also conflict due to the inherent noisiness of neural computation (for discussion, see Knill & Pouget, 2004).

According to the currently influential 'maximum likelihood estimation' (MLE) approach in cognitive science (Bennett et al. 2014 and van Dam et al. 2014 are helpful primers), O-integration occurs when initial estimates of a property provided by different modalities are weighted by their relative reliability and combined in a way that optimizes, i.e., reduces the variance in, the final perceptual estimate of that property.[4] Since this final estimate is a compromise between the different initial estimates (in fact, it will be their weighted average on the MLE approach), O-integration also serves to reduce intersensory discrepancy at the level of conscious perception. 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.

O-integration, as understood here, presupposes that the multisensory causal inference problem has been solved. 'Our brain', as Marc Ernst writes, 'constantly receives sensory information from many different sources and modalities. If corresponding sensory signals are derived from the same object or event, these should be integrated; otherwise, they should be kept separate' (Ernst 2007, 1). Empirical evidence suggests that the extent to which initial estimates of a property across different modalities are O-integrated depends, in part, on the perceptual system's degree of confidence the signals have the same distal source and should be coupled together (Roach et al. 2006, Körding et al. 2007).

Numerous illusions illustrate the perceptual consequences of O-integration. In the *ventriloquism effect* (Bertelson 1999), initial visual and auditory estimates of an object's location conflict. Typically, conscious auditory localization of the object is strongly biased in the direction of the initial visual estimate. When initial estimates are not highly discrepant, the perceptual result is that the sound you *hear* appears (non-veridically) to be coming from the object you *see*. In some of the early literature, this effect is referred to as 'phenomenal fusion' (Radeau & Bertelson 1977). Different views concerning the nature of phenomenal fusion will be discussed in **part two** of this review.

Other well-known illusions exemplifying O-integration include:

- *The McGurk effect*: In this illusion (McGurk & MacDonald 1976), inputs from vision influence the contents of auditory experience. When subjects watch a video of a speaker

articulating the sound /*ga*/ dubbed with a recording of a speaker pronouncing the sound /*ba*/, they report hearing the sound /*da*/ instead (a kind of phonological compromise).

• *Visual capture of touch and proprioception*: When initial visual and haptic estimates of an object's spatial properties, e.g., its shape, size, or orientation, are experimentally set in conflict, the final, conscious haptic estimate is strongly biased in the direction of the initial visual estimate (Gibson 1933, Rock & Harris 1967). Such visual dominance is also found when initial visual and proprioceptive estimates of the position of a body part are made to conflict (Hay et al. 1965; Welch & Warren 1980; Botvinick & Cohen 1998; Samad et al. 2015).

• *The parchment skin illusion*: To elicit this illusion, experimenters recorded the sounds produced while subjects rubbed their palms together (Jousmaki & Hari 1998). These sounds were played back to the participants through headphones, 'dubbing' the tactile stimulation they received. When high frequencies were accentuated, participants reported that their skin felt dry and paper-like. Jousmaki and Hari propose that this illusion reflects an 'omni-present intersensory integration phenomenon, which helps the subject to make accurate tactile decisions about the roughness and stiffness of different textures they manipulate' (R190).

Two points are important. First, ventriloquism and other cases of O-integration can occur *without* the experience of phenomenal fusion. For example, in ventriloquism, you may experience a sound source as closer to a visually perceived object than it really is without thereby being disposed to identify them. This, however, is in keeping with the probabilistic nature of multi-sensory causal inference. O-integration may occur when the probability that multisensory signals have a common distal ground is greater than a certain threshold value, but this value may be lower than that required (in context) for phenomenal fusion to occur.

Second, O-integration seems to be influenced by cognitive factors. Subjects, for example, are more likely to O-integrate visual and auditory signals at large spatial separations when the visual-auditory pairing is familiar than when it is arbitrary (Jackson 1953) and to O-integrate visual and haptic signals when they have the prior, explicit belief that the signals emanate from the same source (Miller 1972, Welch 1972, Helbig & Ernst 2007).[5]

2.3. GENERATIVE MULTISENSORY INTEGRATION

In O-integration, as we have seen, multiple modalities provide different initial estimates of a single property, e.g., a shape, location, or temporal frequency, that they jointly attribute to an object or event. The end-product of O-integration is a revised and, when all goes well, optimized estimate of that property. Hence, the total number of property types represented by different modalities remains the same after O-integration has taken place. What changes is only the way the determinate value of a certain determinable property is represented.

Synthesizing estimates of *different* types of properties across distinct modalities, however, can give rise to the representation of a *new* type of property, one that couldn't be represented by any of the contributing modalities in isolation. In what follows, this 'generative' form of multisensory processing will be referred to as *generative multisensory integration* (or **G-integration** for short).[6]

One example of G-integration is flavor perception. Flavor properties are not represented by any single modality in isolation: they depend instead on the combination of inputs from taste and smell (retronasal olfaction), thermal and somato-sensory cues, as well as information concerning chemical irritation and nociception supplied by the trigeminal system. It is only when these diverse sources of information are G-integrated that a distinctive flavor is perceived,

e.g., the flavor of chocolate or a peaty, single malt Scotch (Auvray & Spence 2008, Spence et al. 2014, Smith 2015).[7]

Other cases of G-integration involve combining *proprioceptive information* about the changing spatial configuration of the body with information derived from other perceptual modalities. I should note that I here use 'proprioception' in a broad way that encompasses the sense of joint position as well as kinesthesis (the sense of bodily movement) and equilibrioception (the sense of balance). Following Macpherson (2011a) and Schwenkler (2013), I also assume that proprioception is properly characterized as a perceptual modality.[8]

The role of proprioception in G-integration can be made concrete by means of two examples:

- *Representing external space:* Multisensory causal inference and, so, O-integration depends on evidence that a single object (or event) is the causal antecedent of stimulation arriving in different modalities. One important kind of bottom-up information for such inference, as we have seen, is *spatial congruence*: signals that arrive from the same region of external space, other things being equal, are more likely to have a common distal cause than those that don't (see Spence 2013, however, for some important restrictions).

The problem, from a neural point of view, is that the exteroceptive senses initially encode spatial information using different, body-relative frames of reference (Stein & Meredith, 1993). An object's location is initially encoded in eye-centered coordinates by vision, hand-centered coordinates by touch, and head-centered coordinates by audition. The receptor surfaces for different senses, moreover, don't stand in invariant spatial relations to one another: the relative position of your eyes, head, and hands change from one moment to the next. How, then, is a stable, crossmodal representation of external space that can support both multisensory causal inference and O-integration achieved?

The solution involves G-integrating modality-specific sources information about an object's location with proprioceptive information about the current spatial configuration of the body. Driver and Spence write:

> When taken in isolation, each modality signals stimulus location with respect to its own receptor surface only (e.g. on the eye for vision, or on the skin for touch). Since the receptors for each modality can move freely relative to external objects (as in eye- or hand-movements) and can also move relative to each other (as when making an eye-movement but no hand-movement, or vice versa), a single modality alone cannot provide a stable representation of external space. Instead, this requires integration of information from multiple sensory modalities, so that current posture is taken into consideration, as well as the location of stimuli on receptor surfaces (Driver & Spence 1998a, 254).

> Generating more useful representations of space (e.g., in order to derive stable visual position across saccades) requires spatial information to be integrated from more than one sense (e.g., retinal position to be combined with extraretinal signals about eye-position). Thus, even when attending to a particular location in just the 'visual' world, it may be that more than one sense is involved in generating a stable representation of that location (Spence & Driver, 2004, 180–81).

Supporting this proposal, single-cell recording studies and models in computational neuroscience indicate that neurons in posterior parietal cortex compute the locations of objects in external space by G-integrating incoming perceptual information with sources of proprioceptive information concerning the spatial configuration of the body (Zipser & Andersen, 1988, Andersen et al. 1997, Cohen & Andersen, 2004, Azañón et al. 2010). For example, the location

of an object relative to the torso can be computed by G-integrating an estimate of the object's location in an eye-centered reference frame with proprioceptive information about the current direction of gaze and the position of the perceiver's head.[9] *Torso-relative location* in external space is thus an example of a new type of property produced by G-integration, a property representable by neither vision nor proprioception alone. The process just described can be iterated by further G-integrating proprioceptive information about the angles of the perceiver's shoulder, elbow, and wrist joints to work up a representation of the object's location relative to her *hand* (another new determinable property). Similar coordinate transformations are possible for the other exteroceptive modalities.

When G-integrated with sources of proprioceptive information, visual, auditory, and tactile signals are thus all capable of representing locations in external space using the same body-relative reference frames (for in-depth, philosophical discussion, see Clark 2011). And, hence, it is possible to discern whether or not the sources of signals arriving from different modalities are spatially congruent, as is often required for both multisensory causal inference and O-integration.

- *Haptic touch:* Unlike passive, cutaneous touch, haptic or 'exploratory' touch enables you to perceive properties such as three-dimensional shape, size, and location in external space. Haptic touch involves both cutaneous stimulations produced by different types of sensors embedded in the skin as well as kinesthetic information about the relative positions of body parts in space (Katz 1925/1989, Gibson 1962, Jones & Lederman 2006, Lederman & Klatzky 2009, Serino & Haggard, 2010). In order to perceive the volumetric, three-dimensional shape of a bottle by means of haptic touch, for example, it is necessary to G-integrate cutaneous information about the bottle delivered by sensors in the skin on your fingertips with up-to-date kinesthetic information about the way your fingers and hands are moving. Haptic touch is a form of G-integration because neither cutaneous touch nor kinesthesis by itself is a source of perceptual information about the voluminous shapes of objects external to your body.

This assessment has been recently challenged by Matthew Fulkerson (2014). According to Fulkerson, multisensory integration is best conceived as an associative relation between two or more unimodal perceptual states that directly assign or bind certain features to an object. For example, in the ventriloquism effect, there is an association between a visual representation as of *an object with certain features at location l_1* and an auditory representation as of *an object with certain features at location l_2* in consequence of which the initial auditory estimate of the object's location is revised in the direction of the initial visual estimate. Haptic touch, Fulkerson suggests, doesn't involve any such associative relation. Instead, '[i]t involves the direct predication or binding of sensory features onto individual objects' (2014, 40). Hence, on this view, haptic touch is unisensory: it doesn't involve what I am referring to as generative multisensory integration.

Let us grant, for purposes of argument, that multisensory integration is well-conceived as an association between two or more unimodally distinct states that directly assign features to objects. Even if we do, the claim that haptic touch doesn't involve multisensory integration is disputable. To perceive a thing's location in external space by means of haptic touch, for example, it is necessary to represent the joint configuration of the mediating limb. This proprioceptive representation, it seems clear, directly assigns certain spatial features to an individual object, namely, your own body, (or a part of your own body). In addition, it is also necessary to form cutaneous representations in response to events that are taking place on the surface of the body, for example, changes in pressure or tension. These cutaneous representations arguably also directly assign certain features to individual objects, whether we identify the latter with the relevant parts of the body, or the thing you are touching, or, as on some accounts, with both (see de Vignemont & Massin, 2015).[10] At the very least, they assign certain somatotopic

locations to the events that elicit them.[11] Similar points can be made with respect to haptic perception of other object properties, such as three-dimensional shape, that aren't representable on the basis of either cutaneous touch or proprioception alone. While this line of argument requires more development than is possible here, it provides initial reason to suppose that haptic touch, contrary to Fulkerson, involves a form of multisensory integration, namely, what I am here referring to as G-integration.

How are O-integration and G-integration related? Processes of G-integration seem to have computational priority inasmuch they support the construction of crossmodally accessible representations of external space. O-integration, as pointed out above, is in many cases constrained by the extent to which signals in different modalities are spatially congruent. It is only by G-integrating proprioceptive information about the configuration of the body with sources of information from the exteroceptive modalities, however, that the brain is able to represent locations in the external world using non-modality-specific spatial reference frames.

2.4. PERCEPTUAL ADAPTATION

Both types of multisensory integration discussed above operate on extremely short time-scales. Perceptual adaptation, by contrast, can be defined as a 'semipermanent change of perception or perceptual-motor coordination that serves to reduce or eliminate a registered discrepancy between or within sensory modalities or the errors in behavior induced by this discrepancy' (Welch 1978, 8). Moreover, in perceptual adaptation, the relevant intersensory discrepancy itself is generally stable and persistent.

One plausible function of perceptual adaptation on long, developmental time-scales is to prevent intersensory discrepancies in spatial localization and/or errors in sensorimotor control due to biologically normal changes in body metrics (Held 1965; Redding & Wallace 1997, chap. 2; Bedford 1999). These changes in body metrics must be taken into account if different modalities are successfully to localize objects in external space using a common set of body-relative reference frames. Consider what is required in order to represent an object's location using the right hand (this is an example of haptic touch, discussed in section 2.3 above). First, it is necessary to have accurate proprioceptive information about the current spatial configuration of the body, in particular, information about the joint angles, respectively, assumed by the shoulder, elbow, and wrist. Such information is represented by what Brian O'Shaughnessy (1980) refers to as the *short-term body image*.[12] It is also necessary, however, to have up-to-date information concerning the comparatively stable shapes and sizes of the limb segments connected at those joints. These and other non-momentary bodily properties are represented by what O'Shaughnessy calls the *long-term body image* and by what cognitive scientists sometimes refer to as a stored 'body model' (Tsakiris 2010, Longo and Haggard, 2010). If, however, the latter type of information is inaccurate because a child's limbs have grown longer or her shoulders have grown broader, then the location assigned to the object by touch may conflict with the location assigned to it by vision. Such intersensory discrepancy can be resolved by perceptual adaption processes that function to up-date internal representations of the body's metric properties contained in the long-term body image.

Most conclusions about the nature of perceptual adaptation have been based on studies of the effects of devices that reverse, displace, or otherwise distort the retinal image, thereby causing spatial information available by means of vision to conflict with information in other modalities. In what follows, I shall refer to these as optical rearrangement devices (or **ORDs** for short).

Some of the earliest experiments on optical rearrangement were performed by Helmholtz (1924/2005), who practiced reaching to targets while wearing prisms that displaced the retinal image 16–18° to the left (also see Stratton 1897, 1899). The initial tendency was to reach too far in the direction of lateral displacement, but after a number of trials reaching gradually regained

its former level of accuracy. Further, Helmholtz discovered that immediately after removing the prisms from his eyes errors in were made in the opposite direction, i.e., when reaching for a target, he now moved his hand too far to the right. This *negative after-effect* is standardly used in current experimentation as a measure of the extent to which perceptual adaptation to lateral displacement has occurred.

Helmholtz's experiments catalyzed a research tradition on ORD adaptation that experienced its heyday for two decades between 1960 and 1980 (for helpful guides, see Welch 1978 and Redding & Wallace 1997). Two questions dominated studies conducted during this period. First, what are the necessary and sufficient conditions for adaptation to occur? Second, just what happens when subjects adapt to perceptual rearrangement? What is the *end product* of the relevant form of perceptual learning?

It was suggested above that adaptation to developmental changes in body metrics occurs at the level of the proprioceptive, long–term body image. By contrast, numerous studies indicate that a significant amount of adaptation to ORDs occurs rather at the level of the *short-term* body image (as originally proposed by Stratton 1897, 1899). On this view, the end product of adap–tation to an ORD is a recalibration of position sense at one or more points of articulation in the body. As you practice reaching for a target while wearing laterally displacing prisms, for exam–ple, the muscle spindles, joint receptors, and Golgi tendon organs in your shoulder and arm con–tinue to generate the same patterns of nerve impulses (action potentials) as before, but the *meaning* assigned to those patterns by the brain undergoes change: whereas before they signified that your arm was moving along one path through the seven-dimensional space of possible arm configurations, they gradually come to signify that it is moving along a different path in that kinematic space, namely, the one consistent with the prismatically distorted visual feedback you are receiving.[13] Similar recalibrations of the short-term body image are also possible with respect to sources of information for head and eye position. After adapting to laterally displacing prisms, signals from receptors in your neck that previously signified the alignment of your head and bodily midline, for example, may come to signify that your head is turned slightly to the side (for discussion, see Harris 1965, 1980 and Welch 1978, chap. 3).

2.5. CROSSMODAL SPATIAL ATTENTION

Numerous studies have found that attending to a region of space in one perceptual modality tends to attract attention to that region in other modalities (for overviews, see Spence 2010, 2014). Hearing an unanticipated sound to one's right, for example, can enhance the speed and accuracy with which targets subsequently presented on the right are discriminated by vision and touch. Such crossmodal facilitation has also been found to obtain when attention is shifted toward a cue endogenously, i.e., in a voluntary, top-down manner.

Driver and Spence's studies indicate that the crossmodal links in selective attention involve representations of cue and target locations in *external space*. What matters, for example, in tactile-visual attentional facilitation isn't simply whether cutaneous stimulation is delivered to the skin on the left or right side of your body and, so, whether the tactile cue is registered by the left or right hemisphere of your brain. What matters rather is whether the stimulated region of the bodily surface is located to the left or right of the direction of gaze at the time of stimu–lation, even if that region is not currently visible (Kennett et al. 2002). Parallel points hold with respect to visual-tactile facilitation: when you are looking straight ahead with your arms crossed in front of you, a flash on the right side of your visual field will advantage tactile discriminations with your left hand, and vice-versa.

These results reveal that the spatial mapping between attentionally selected locations in different modalities is updated as the proprioceptively represented positions of the eyes, hands,

other parts of the body change. 'A third modality (here proprioception)', as Driver and Spence write, 'can apparently influence the attentional interactions between two other modalities' (1998b, 1322). To use the terminology of section 2.3, proprioceptive information is computationally merged with relevant sources of non-proprioceptive, perceptual information by means of G-integration in order to represent the locations of cues and targets in external space, and it is toward locations within this space that attention across different modalities is directed.

2.6. SUMMARY: A CARTOGRAPHY OF MULTISENSORY PROCESSES

It may be useful, in concluding, to highlight some of the computational dependency relations between the various forms of multisensory processing discussed above (Fig. 1).

First, *perceptual adaptation* (section 2.4) to intersensory discrepancies arising from changes in body metrics in the course of normal development, it was suggested, functions to maintain the accuracy of the proprioceptive, long-term body image. An up-to-date long-term body image, in turn, is necessary effectively to combine perceptual information in modality-specific reference frames with proprioceptive information by means of *G-integration* (section 2.3). Since such perceptual–proprioceptive G-integration is required to represent locations in external space crossmodally (Driver & Spence 1998a/b, Clark 2011), it is also required for the operation of *crossmodal spatial attention* (section 2.5) and, in addition, for the capacity to discern whether multisensory signals are spatially congruent. As pointed out above, spatial congruence is an important source of bottom-up, perceptual information used in *multisensory causal inference* (section 2.1). Hence, the latter is also computationally dependent, in part, on crossmodally accessible representations of external space that arise from *G-integration*. Finally, solving the multisensory causal inference problem is required for *O-integration* to occur (section 2.2). Different initial estimates of a property, for example, shape, location, or temporal frequency, should be O-integrated if they pertain to the same object (or event) and segregated if they do not.[14]
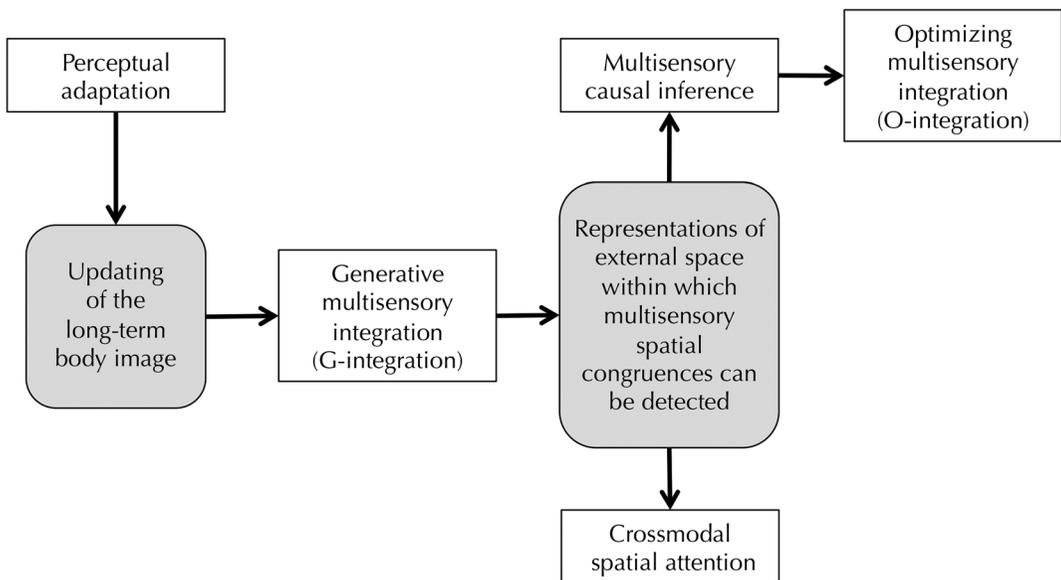

Figure 1. Some of the important functional relationships between different multisensory processes.

Acknowledgements

Short Biography

Robert Briscoe is Associate Professor of philosophy at Ohio University. His research deals with topics in the philosophy and cognitive science of perception as well as the philosophy of mind. Special interests include the role of action in perception, spatial representation, mental imagery, depiction, and pictorial experience. Prior to his appointment at Ohio University, he was a research associate at the MIT Initiative for Technology and Self and taught in the philosophy department at Loyola University in New Orleans. He has a BA from Columbia University and a PhD from Boston University.

Notes

* Correspondence: Department of Philosophy, Ohio University, 202 Ellis Hall, Athens, OH 45701, USA. Email: rbriscoe@gmail.com

[1] Helpful overviews can be found in Stein & Meredith (1993), Calvert et al. (2004), Spence & Driver (2004), Trommershäuser et al. (2011), Stein (2012), and Bennett & Hill (2014). Stokes et al. (2014) and Matthen (2015) contain essays that address philosophical implications of multisensory processing and debates concerning multisensory consciousness.

[2] This general feature distinguishes the multisensory processes reviewed here from other interactions between the senses, such as synaesthesia, that do not have a perception-enhancing function. For discussion, see O'Callaghan 2012.

[3] See Clark (2013) and Rescorla (2015) for insightful, philosophically oriented discussions.

[4] This is a characterization of multisensory integration at the perceptual, information-processing level. At the neurophysiological level, multisensory integration is defined as occurring when a 'neural signal is different (e.g., bigger, smaller, having a different temporal evolution) than each of the presumptive component responses and cannot readily be deconstructed to yield the unique contributions to its formation' (Stein et al. 2010, 1717).

[5] For skeptical assessments, see Bertelson et al. (2001) and Deroy (2014).

[6] Thanks to Bence Nanay for a helpful discussion.

[7] In the highly ramified taxonomy of multisensory experiences recently proposed by Fiona Macpherson, the conscious upshot of flavor processing, so understood, is a 'cross-modal$_{within-sensory-integration-novel}$ experience' (Macpherson 2011b, 449–50). In such experience, there is 'some brand new content' over and above the contents of the contributing modalities.

[8] But for opposing views, see Gallagher (2005) and Bermúdez (2005).

[9] Sources of information about direction of gaze include both afference from the extraocular muscles and efference copy, i.e., a copy of the 'outflowing' motor command to execute an eye movement. For discussion of their relative importance in visual space perception, see Bridgeman (2010).

[10] de Vignemont and Massin (2015), following Weber, defend a pressure theory of cutaneous touch. On their account, cutaneous touch is the direct perception of a 'a symmetrical relation: we feel something pressing on this part of our body (or that this part of our body is pressing on something)…. [I]ts proper and primary objects are relations, whose relata are generally parts of our body and external objects'. Hence, cutaneous touch typically assigns features both to the thing exerting pressing on a region R of your bodily surface and to R at the same time. I'm grateful to Matthew Longo and Frédérique de Vignemont for discussion of this view.

[11] There is also evidence that cutaneous touch individuates objects and events in accord with Gestalt grouping principles (Kitagawa et al. 2009, Gallace & Spence 2011).

[12] Other authors refer alternatively to the *postural schema* (Head & Holmes 1911, Longo and Haggard, 2010) and the *body schema* (Gallagher 1986, 2005).

[13] The human arm has seven degrees of freedom: three at the wrist, one at the elbow, and three at the shoulder.

[14] A more expansive cartography, I should mention, would deal with the role of O-integration in various types of bodily experiences (see, e.g., Maravita et al. 2003, Petkova et al. 2011, de Vignemont 2014b). Also omitted from discussion in this review is a form of visual-somatosensory interaction that Haggard et al. (2013) call 'multimodal modulation'. In

multimodal modulation, seeing the body affects somatosensory processing, even when vision is entirely non-informative about the somatosensory stimulation being received. Viewing one's arm, for example, enhances the spatial acuity of cutaneous touch even when the arm is subsequently hidden from sight at the moment tactile stimulation is delivered (Kennett et al. 2001). Non-informative vision of the body has also been found to decrease the perceived intensity of acute pain (Longo et al. 2009) and to shrink the perceived size of objects touching the skin (Longo & Sadibolova 2013).

Works Cited

Andersen, R., L. Snyder, D. Bradley, and J. Xing. 'Multimodal Representation of Space in the Posterior Parietal Cortex and Its Use in Planning Movements.' *Annual Review of Neuroscience* 20.1 (1997): 303–330.

Auvray, M., and C. Spence. 'The Multisensory Perception of Flavor.' *Consciousness and Cognition* 17.3 (2008): 1016–1031.

Azañón, E., M. Longo, S. Soto-Faraco, and P. Haggard. 'The Posterior Parietal Cortex Remaps Touch into External Space.' *Current Biology* 20.14 (2010): 1304–1309.

Bayne, T. 'The Multisensory Nature of Perceptual Consciousness.' *Sensory Integration and the Unity of Consciousness*. Eds. D. Bennett and C. Hill. Cambridge, MA: MIT Press 2014. 15–36.

Bedford, F. 'Keeping Perception Accurate.' *Trends in Cognitive Sciences* 3.1 (1999): 4–11.

——. 'Towards a General Law of Numerical/Object Identity.' *Current Psychology of Cognition* 20.3/4 (2001): 113–176.

Bennett, D., and C. Hill. *Sensory Integration and the Unity of Consciousness*. Cambridge, MA: MIT Press, 2014.

Bennett, D., J. Trommershäuser, and L. van Dam. 'Bayesian Modeling of Perceiving: A Guide to Basic Principles.' *Sensory Integration and the Unity of Consciousness*. Eds. D. Bennett and C. Hill. Cambridge, MA: MIT Press, 2014. 3–14.

Bermúdez, J. 'The Phenomenology of Bodily Awareness.' *Phenomenology and Philosophy of Mind*. Eds. D. Woodruff Smith and A. Thomasson. Oxford: Clarendon Press, 2005. 295–316.

Bertelson, P. 'Ventriloquism: A Case of Crossmodal Perceptual Grouping.' *Advances in Psychology* 129 (1999): 347–362.

Bertelson, P., J. Vroomen, G. Aschersleben, and B. de Gelder. 'Object Identity Decisions: At What Processing Levels? Or: Why the Cantaloupe Might Work.' *Current Psychology of Cognition* 20.3/4 (2001): 177–182.

Botvinick, M., and J. Cohen. 'Rubber Hands 'Feel' Touch That Eyes See.' *Nature* 391.6669 (1998): 756–756.

Bridgeman, B. 'Space Constancy: The Rise and Fall of Perceptual Compensation. *Space and Time in Perception and Action*. Eds. R. Nijhawan and B. Khurana. Cambridge: Cambridge University Press, 2010. 94–108.

Calvert, G., C. Spence, and B. Stein, eds. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press, 2004.

Clark, A. 'Cross Modal Links and Selective Attention.' *The Senses: Classic and Contemporary Philosophical Perspectives*. Ed. Fiona MacPherson. Oxford: Oxford University Press, 2011. 375–396.

——. 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.' *Behavioral and Brain Sciences* 36.03 (2013): 181–204.

Cohen, Y., and R. Andersen. 'Multisensory Representations of Space in the Posterior Parietal Cortex.' *The Handbook of Multisensory Processes*. Eds. C. Spence Calvert and B. Stein. Cambridge, MA: MIT Press, 2004. 463–479.

de Vignemont, Frédérique. 'A Multimodal Conception of Bodily Awareness.' *Mind* 123.492 (2014a): 989–1020.

——. 'Multimodal Unity and Multimodal Binding.' *Sensory Integration and the Unity of Consciousness*. Eds. D. Bennett and C. Hill. Cambridge, MA: MIT Press, 2014b. 125–150.

de Vignemont, F., and O. Massin. 'Touch.' *The Oxford Handbook of the Philosophy of Perception*. Ed. Mohan Matthen. Oxford: Oxford University Press, 2015.

Deroy, O. 'The Unity Assumption and the Many Unities of Consciousness.' *Sensory Integration and the Unity of Consciousness*. Eds. D. Bennett and C. Hill. Cambridge, MA: MIT Press, 2014. 105–124.

Driver, J., and C. Spence. 'Attention and the Crossmodal Construction of Space.' *Trends in Cognitive Sciences* 2.7 (1998a): 254–262.

——. 'Cross–Modal Links in Spatial Attention.' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 353.1373 (1998b): 1319–1331.

Ernst, M. 'A Bayesian View on Multimodal Cue Integration.' *Human Body Perception from the Inside Out*. Eds. G. Knoblich, I. M. Thornton, M. Grosjean, and M. Shiffrar. New York, NY: Oxford University Press, 2006. 105–131.

——. 'Learning to Integrate Arbitrary Signals from Vision and Touch.' *Journal of Vision* 7.5 (2007): 1–14.

Fulkerson, M. *The First Sense: A Philosophical Study of Human Touch*. Cambridge, MA: MIT Press, 2014.

Gallace, A., and C. Spence. 'To What Extent Do Gestalt Grouping Principles Influence Tactile Perception?' *Psychological Bulletin* 137 (2011): 538–561.

Gallagher, S. 'Body Image and Body Schema: A Conceptual Clarification.' *Journal of Mind and Behavior* 7 (1986): 541–554.

——. *How the Body Shapes the Mind*. Oxford: Clarendon Press, 2005.

Gibson, J. J. 'Adaptation, After-Effect and Contrast in the Perception of Curved Lines.' *Journal of Experimental Psychology* 16 (1933): 1–31.

——. 'Observations on Active Touch.' *Psychological Review* 69.6 (1962): 477–491.

Haggard, P., G. Iannetti, and M. Longo. 'Spatial Sensory Organization and Body Representation in Pain Perception.' *Current Biology* 23.4 (2013): R164–R176.

Harris, C.S. 'Perceptual Adaptation to Inverted, Reversed, and Displaced Vision.' *Psychological Review* 72.6 (1965): 419–444.

——. 'Insight or Out of Sight? Two Examples of Perceptual Plasticity in the Human Adult.' *Visual Coding and Adaptability*. Ed. C.S. Harris. Hillsdale, NJ: Lawrence Erlbaum, 1980. 95–149.

Hay, J., H. Pick, and K. Ikeda. 'Visual Capture Produced by Prism Spectacles. *Psychonomic Science* 2 (1965): 215–216.

Head, H., and G. Holmes. 'Sensory Disturbances from Cerebral Lesions.' *Brain* 34.2-3 (1911): 102–254.

Helbig, H., and M. Ernst. 'Knowledge about a Common Source Can Promote Visual-Haptic Integration.' *Perception* 36.10 (2007): 1523–1534.

Held, R. 'Plasticity in Sensory-Motor Systems.' *Scientific American* 213.5 (1965): 84–94.

Jackson, C. 'Visual Factors in Auditory Localization.' *Quarterly Journal of Experimental Psychology* 5 (1953): 52–65.

Jones, L., and S. Lederman. *Human Hand Function*. Oxford: Oxford University Press, 2006.

Jousmäki, V., and R. Hari. 'Parchment-Skin Illusion: Sound–Biased Touch.' *Current Biology* 8.6 (1998): R190–R191.

Katz, D. *The World of Touch*. Trans. L. Krueger. Hillsdale, NJ: Lawrence Erlbaum, 1925/1989.

Kennett, S., M. Taylor-Clarke, and P. Haggard. 'Noninformative Vision Improves the Spatial Resolution of Touch in Humans.' *Current Biology* 11 (2001): 1188–1191.

Kennett, S., C. Spence, and J. Driver. 'Visuo-Tactile Links in Covert Exogenous Spatial Attention Remap across Changes in Unseen Hand Posture.' *Perception & Psychophysics* 64 (2002): 1083–1094.

Kitagawa, N., Y. Igarashi, and M. Kashino. 'The Tactile Continuity Illusion.' *Journal of Experimental Psychology: Human Perception and Performance* 35.6 (2009): 1784–790.

Knill, D., and A. Pouget. 'The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation.' *Trends in Neurosciences* 27.12 (2004): 712–719.

Körding, K., U. Beierholm, W. Ma, S. Quartz, J. Tenenbaum, and L. Shams. 'Causal Inference in Multisensory Perception.' *PloS One* 2.9 (2007): e943.

Landy, M., M. Banks, and D. Knill. 'Ideal-Observer Models of Cue Integration.' *Sensory Cue Integration*. Eds. J. Trommershäuser, K. Kording, and M. Landy. Oxford: Oxford University Press, 2011. 5–29.

Lederman, S., and R. Klatzky. 'Haptic Perception: A Tutorial.' *Attention, Perception, & Psychophysics* 71.7 (2009): 1439–1459.

Longo, M., *et al.* 'Visually Induced Analgesia: Seeing the Body Reduces Pain.' *The Journal of Neuroscience* 29.39 (2009): 12125–12130.

Longo, M., and P. Haggard. 'An Implicit Body Representation Underlying Human Position Sense.' *Proceedings of the National Academy of Sciences* 107 (2010): 11727–11732.

Longo, M., and R. Sadibolova. 'Seeing the Body Distorts Tactile Size Perception.' *Cognition* 126.3 (2013): 475–481.

Macpherson, F., ed. *The Senses: Classic and Contemporary Philosophical Perspectives*. Oxford: Oxford University Press, 2011a.

Macpherson, F. 'Cross Modal Experiences.' *Proceedings of the Aristotelian Society* 111.3 (2011b): 429–468.

Maravita, A., C. Spence, and J. Driver. 'Multisensory Integration and the Body Schema: Close to Hand and Within Reach.' *Current Biology* 13.13 (2003): R531–R539.

Matthen, M., ed. *The Oxford Handbook of the Philosophy of Perception*. Oxford: Oxford University Press, 2015.

McGurk, H., and J. MacDonald. 'Hearing Lips and Seeing Voices.' *Nature* 264 (1976): 746–748.

Miller, E. 'Interaction of Vision and Touch in Conflict and Nonconflict Form Perception Tasks.' *Journal of Experimental Psychology* 96 (1972): 114–123.

O'Callaghan, C. 'Perception and Multimodality.' *The Oxford Handbook of Philosophy of Cognitive Science*. Eds. E. Margolis, R. Samuels, and S. Stich. Oxford: Oxford University Press (2012): 92–117.

——. 'Not All Perceptual Experience is Modality Specific.' *Perception and Its Modalities*. Eds. D. Stokes, M. Matthen, and S. Biggs. Oxford: Oxford University Press, 2014: 133–165.

O'Shaughnessy, B. *The Will*. 1 vol. Cambridge: Cambridge University Press, 1980.

Petkova, V., *et al.* 'From Part to Whole-Body Ownership in the Multisensory Brain.' *Current Biology* 21.13 (2011): 1118–1122.

Radeau, M., and P. Bertelson. 'Adaptation to Auditory-Visual Discordance and Ventriloquism in Semirealistic Situations.' *Perception & Psychophysics* 22 (1977): 137–146.

Redding, G., and B. Wallace. *Adaptive Spatial Alignment*. Mahwah, NJ: Erlbaum, 1997.

Rescorla, M. 'Bayesian Perceptual Psychology.' *The Oxford Handbook of the Philosophy of Perception*. Ed. Mohan Matthen. forthcoming 2015.

Roach, N., J. Heron, and P. McGraw. 'Resolving Multisensory Conflict: A Strategy for Balancing the Costs and Benefits of Audio–Visual Integration.' *Proceedings of the Royal Society B: Biological Sciences* 273.1598 (2006): 2159–2168.

Robertson, L. 'Binding, Spatial Attention and Perceptual Awareness.' *Nature Reviews Neuroscience* 4.2 (2003): 93–102.

Rock, I., and C.S. Harris. 'Vision and Touch.' *Scientific American* 216 (1967): 96–107.

Samad, M., A. J. Chung, and L. Shams. 'Perception of Body Ownership Is Driven by Bayesian Sensory Inference.' *PloS one* 10.2 (2015): e0117178.

Schwenkler, J. 'The Objects of Bodily Awareness.' *Philosophical Studies* 162.2 (2013): 465–472.

Serino, A., and P. Haggard. 'Touch and the Body.' *Neuroscience & Biobehavioral Reviews* 34.2 (2010): 224–236.

Shams, L. 'Early Integration and Bayesian Causal Inference in Multisensory Perception.' *The Neural Bases of Multisensory Processes.* Eds. M. Murray and M. Wallace. Boca Raton, FL: CRC Press, 2012: 217–231.

Shams, L., and U. Beierholm. 'Causal Inference in Perception.' *Trends in Cognitive Sciences* 14.9 (2010): 425–432.

Smith, B. 'The Chemical Senses.' *The Oxford Handbook of the Philosophy of Perception.* Ed. Mohan Matthen. Oxford: Oxford University Press, 2015.

Spence, C., and J. Driver, eds. *Crossmodal Space and Crossmodal Attention.* Oxford: Oxford University Press, 2004.

Spence, C. 'Crossmodal Spatial Attention.' *Annals of the New York Academy of Sciences* 1191 (2010): 182–200.

——. 'Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule.' *Annals of the New York Academy of Sciences* 1296 (2013): 31–49.

——. 'Orienting Attention: A Crossmodal Perspective.' *The Oxford Handbook of Attention* Eds. K. Nobre, A. Nobre, and S. Kastner. Oxford: Oxford University Press, (2014): 446–471.

Spence, C., M. Auvray, and B. Smith. 'Confusing Tastes with Flavours.' *Perception and Its Modalities.* Eds. D. Stokes, M. Matthen, and S. Biggs. Oxford: Oxford University Press, 2014: 247–276.

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

Stein, B., ed. *The New Handbook of Multisensory Processing.* Cambridge, MA: MIT Press, 2012.

——. *et al.* 'Semantic Confusion Regarding the Development of Multisensory Integration: A Practical Solution.' *European Journal of Neuroscience* 31.10 (2010): 1713–1720.

Stokes, D., M. Matten, and S. Biggs, eds. *Perception and Its Modalities.* Oxford: Oxford University Press, 2014.

Stratton, G.M. 'Vision Without Inversion of the Retinal Image.' *Psychological Review* 4.5 (1897): 341–360, 463–481.

——. 'The Spatial Harmony of Touch and Sight.' *Mind* 8 (1899): 492–505.

Tsakiris, M. 'My Body in the Brain: A Neurocognitive Model of Body-Ownership.' *Neuropsychologia* 48.3 (2010): 703–712.

Treisman, A. 'Solutions to the Binding Problem: Progress through Controversy and Convergence.' *Neuron* 24.1 (1999): 105–125.

——. 'Consciousness and Perceptual Binding.' *The Unity of Consciousness: Binding, Integration, Dissociation.* Ed. A. Cleeremans. Oxford England: Oxford University Press (2003). 95–113.

Trommershäuser, J., K. Kording, and M. Landy, eds. *Sensory Cue Integration.* Oxford: Oxford University Press, 2011.

van Dam, L., C. Parise, and M. Ernst. 'Modeling Multisensory Integration.' *Sensory Integration and the Unity of Consciousness.* Eds. D. Bennett and C. Hill. Cambridge, MA: MIT Press, 2014. 209–230.

von Helmholtz, H. *Treatise on Physiological Optics.* 3 vols. Mineola, NY: Dover Publications, 1924/2005.

Welch, R. 'The Effect of Experienced Limb Identity upon Adaptation to Simulated Displacement of the Visual Field.' *Perception and Psychophysics* 12 (1972): 453–456.

——. *Perceptual Modification: Adapting to Altered Sensory Environments.* New York: Academic Press, 1978.

——. 'Meaning, Attention, and the "Unity Assumption" in the Intersensory Bias of Spatial and Temporal Perceptions.' *Advances in Psychology* 129 (1999): 371–387.

Welch, R., and D. Warren. 'Immediate Perceptual Response to Intersensory Discrepancy.' *Psychological Bulletin* 88.3 (1980): 638–667.

Zipser, D., and R. Andersen. 'A Back Propagation Programmed Network that Simulates the Response Properties of a Subset of Posterior Parietal Neurones.' *Nature* 331 (1988): 679–684.