Perception and Multimodality

Casey O'Callaghan

Oxford Handbook of Philosophy and Cognitive Science

eds. Margolis, Samuels, and Stich

27 October 2008

1 Unimodal approaches to the study of perception

Sometimes you turn to look at the source of a *clunk*. Sometimes you see a pan on the stove and grasp its hot handle. Sometimes you smell the pepper you taste, or find a stinky dead mouse under the sink. In each case, you perceive using more than one sense. You hear the sound and see its source. You feel and see the handle. You smell and taste the pepper. You see and smell the mouse. By itself, this is no surprise: we perceive with vision, audition, gustation, olfaction, and proprioception.

And yet, most work on perception by philosophers and cognitive scientists has focused on vision. Empirical psychologists catalog visual illusions and deficits because they want to discover perceptual principles and mechanisms. Philosophers analyze visual experience to explain what we perceive and how we perceive it. This focus makes sense since vision is central to how most humans experience, understand, and navigate their surroundings.

Focusing exclusively on vision is risky. Nothing guarantees that what holds of seeing holds generally of perceiving. What is essential to seeing might not be essential to perceiving, and certain requirements on perceiving might lack salience in the visual case. Impressive diversity among the varieties of perceiving also might go underappreciated. Attention to the contrasts among smelling a scallion's odor, hearing the sounds of a colleague's footsteps, and seeing a building's collapse tells us things about perception that vision cannot. Focusing upon modalities other than vision is good methodology in the study of perception (see, e.g, Matthen, 2005; Clark, 2006).

Researchers in recent years increasingly have turned to nonvisual modalities for insights about perception and perceptual processes. Audition has attracted interest among scientists and philosophers, who have contrasted the objects and the spatial and temporal aspects of audition and vision (see, e.g., Bregman, 1990; Handel, 1993; McAdams and Bigand, 1993; Blauert, 1997; Casati and Dokic, 1994; O'Callaghan, 2007, 2008a). Others have explored smell and taste in order to test the claim that perceptual phenomenology is a variety of representational content (see Lycan, 2000; Batty, 2007). Martin (1992) contrasts tactile spatial experience with visual spatial experience and expresses skepticism whether any general account of perception holds across modalities. Noë (2004) even argues that visual awareness essentially involves the kind of exploratory activity characteristic to tactile and proprioceptive perception (see also O'Regan and Noë, 2001). Promising interdisciplinary research on perceptual modalities other than vision now is beginning to thrive. Such work raises new puzzles and invites novel questions, and it represents the kind of multimodal approach necessary to avoid a parochial, vision-based take on perception.

This approach nonetheless risks remaining *unimodal* in one striking way. It involves investigating each modality in isolation from the others. It is easy to find examples from empirical and philosophical literature. Sensory scientists study the transduction and transmission of information in a single sense or sensory pathway, or study the responses of cells or networks of cells to unimodal stimuli like flashes, beeps, odors, or pinpricks. Philosophers of perception write about unimodal qualia or sensible qualities such as hue, pitch, and taste. When the topic is not unimodal characteristics, it frequently is unimodal experiences such as the visual experience of seeing a shape, the auditory experience of hearing a crash, or the olfactory experience of smelling a corpse flower. As a result, theorizing about perception often takes place like this: we think about individual modalities, compare and contrast them, and compile the stories into an overall account of what it is to perceive.

This approach reveals a presupposition that is not innocent. The methodology suggests that we are thinking of each sense as *explanatorily independent* from the others. This assumption of independence is evident at physiological, functional, and experiential levels of explanation. At one extreme, the senses on this way of thinking involve different organs and pathways, perform discrete functions, are causally and informationally encapsulated, and constitute entirely distinctive modes of awareness. The modalities of sense perception, and their associated sense experiences, thus, in principle, are separable from each other. It is tempting to assume, when examining just one sense, that it is possible to get a complete account of perceiving with that sense modality that is explanatorily independent from the others. If so, the senses constitute independent domains of philosophical and scientific inquiry. One might even hold that assembling the sense-specific stories about seeing, hearing, touching, smelling, and tasting exhausts theorizing about perception.

What supports this approach? I suspect it is the conviction that the modalities of perception and perceptual awareness *differ* so dramatically. Distinctive sense organs respond to different forms of energy. These organs function in quite different ways, and each seems dedicated to performing a specialized kind of task. Distinct pathways of activation and areas of the brain are dedicated to the different senses. Experiences you associate with different modalities also seem to differ remarkably. Seeing a pileated woodpecker differs in qualitatively dramatic ways from hearing it vocalize. Smelling unfamiliar odors and hearing musical sounds make it intuitive to think of each sense as experientially distinctive. Furthermore, the modalities differ in their effects. Each grounds different sorts of beliefs and actions. When riding your bike, thanks to vision, you come to believe a pedestrian is crossing the street ahead; thanks to hearing, you get out of the way of the car approaching from behind.

This perspective, and the methodology, would be viable if experiences associated with the different senses were entirely experientially discrete or encapsulated, or if they were exhausted by distinctive or sense-specific phenomenology. It would be defensible if the senses had disjointed functions, or if they were entirely causally independent or physiologically isolated.

The evidence, however, suggests that we should abandon this perspective. Perceiving is *richly* multimodal. Perceptual processes involve extensive interaction among sensory modalities. The patterns of interaction demonstrate significant functional cooperation among the senses. Adequately accounting for multimodal processes—those that involve interaction or coordination among more than one sense—suggests that explaining what happens in one modality requires appeal to others. The senses thus are not causally, functionally, or explanatorily independent. We may see and hear, but what we see depends upon what we hear. I will suggest that this extends even to perceptual experience, where the products of such cooperation are evident in multimodal perceptual content and phenomenology. Any adequate philosophical understanding of perception and perceptual content should explain the respects in which perception is multimodal.

Attention to the multimodal aspects of perception challenges common assumptions about the content and phenomenology of perception, and about the individuation and psychological nature of sense modalities. Multimodal perception thus presents a valuable opportunity for a case study in mature interdisciplinary cognitive science. This chapter aims to raise these issues against the historical background of unimodal approaches in the study of perception. It presents some of the central empirical findings concerning multimodality, and it explains the philosophical implications of these findings. Foremost, it aims to encourage and open avenues for future research.

2 Some varieties of multimodality

We should from the start distinguish several ways in which perception could be multimodal. Though not exhaustive, this will help to highlight the targets.

First, we perceive thanks to *several* senses. Intuitively, we see, hear, touch, taste, and smell. The commonsense distinctions among sense modalities might correspond to folk psychological categories used to explain different ways of acquiring empirical beliefs or different patterns of responding to some surroundings. Seeing makes you believe that the car across the street is purple, and hearing makes you turn your head toward the collision. These categories perhaps are etiologically grounded in differences to perceptual experience that are recognizable from the first person—nobody would mistake a current experience of seeing a purple car for hearing. But a great deal of science also has taught us in fascinating detail about the physiological differences among the eyes, ears, skin, tongue, and nose. It has taught us about nerves and brain activity, and about the evolution and function of sensory organs and pathways. It is not trivial to say that we perceive with different sense modalities, but it is a weak claim that in this respect alone perception is multimodal.

Despite their differences, the senses work in concert. Sensing frequently occurs in a multimodal context, among sources that collectively or individually stimulate more than one modality. Sensory organs and pathways are not activated entirely in isolation.

Activity in the different senses is responsible for your overall perceptual experience at any given time. Right now, I hear some music and feel a breeze, see a sandwich and taste a sour candy. But although perceptual experience comprises experiences drawn from the different senses, it does not seem fragmented or disjointed. There is a strong sense in which one's visual experience, auditory experience, tactile experience, and so on, make up a single composite experience with visual, auditory, tactile, and so on, aspects or components. In this respect, perceptual experience is multimodally *unified* (see, e.g., Tye, 2003; Bayne and Chalmers, 2003).

Certain experiences, however, seem to belong together in a stronger sense, as when you both hold and see a fuzzy yellow tennis ball, or when you reach up to a location in space to grab a tennis ball you see approaching. Or when you taste something you smell. In these cases, sets of experiences associated with different sense modalities are *integrated* in a way that other unified experiences, such as concurrent experiences of tasting some wine and seeing some flowers, are not. They are not merely co-conscious and do not simply constitute a conjoint experience. Experiences with quite different subject matter might be unified in the sense above, but integration involves experiences that concern a common perceptible object or feature. Integration is unification of experiences that are the same with respect to what is perceived, rather than just unification among experiences that concern different things (integration thus is closely related to what Bayne and Chalmers (2003) call *objectual unity*).

A natural question, given the dramatic differences among experiences associated with different senses, is how they could seem to converge upon the same items or features. Empiricists have held that identification of common objects of sensory experience requires extra-perceptual cognition, such as inference or association. For this reason, empiricists traditionally answered Molyneux's question negatively. A traditional empiricist thus would attribute what I have called integration to processes beyond sense perception. One lesson of this chapter is that we should abandon the traditional view. While much sensory processing may occur in distinctive sensory pathways, information from multiple modalities commonly is assembled, knit together, and integrated into a single, unified experience. Explaining the perceptual mechanisms responsible for unification and integration would be a great advance in the cognitive science of perception, and it would impact a number of philosophical debates about perception.

An intriguing class of cases suggests a sense in which perceptual *processes* themselves are multimodal. These cases reinforce the cooperative character of perceptual modalities, and ultimately help to explain multimodal unification and integration.

First, consider *adaptation*. When stimulation in a given modality is disrupted in order to introduce a discrepancy between experiences in two modalities, subjects adapt over time. For example, despite being disoriented when fitted with prism goggles that shift the optical scene, subjects later adapt and manage to get around normally (Helmholtz, 1866/1925; Held, 1965). Adaptation involves adjustments to perceptual experience over time. Such reordering demonstrates a tendency to *calibrate* distinct modalities to each other, and adaptation illustrates a concern for the *relationships* among sense modalities. Adaptive processes are multimodal in that they reorganize experience in light of information drawn from multiple senses.

Someone might say that adaptive processes are not strictly speaking perceptual, but merely reorganize perceptual information for the purposes of action. Adaptation might involve adjustments to perceptual experience over time, but it also might just amount to the capacity to cope with disrupted perceptual stimulation in the face of discrepant experience.

A more dramatic class of cases demonstrates that the multimodality of sense perception runs deeper. These involve *interaction* among sense modalities. Frequently, one modality impacts processes and alters experiences associated with another. Interactions among the senses are rampant, and they commonly even lead to perceptual *illusions*. To be clear, stimulation to one modality causes perceptual illusions in another. Seeing can make you have illusory tactile feelings, and hearing can cause visual illusions. This is not just adaptation. Adaptation results from the ability of perception or action systems to adjust over time. It may have no immediate and apparent effect, and it persists after the intervention. Intermodal interactions and illusions happen immediately, and their effects disappear as soon as the stimulus does. The effects of adaptation and interaction might even compete, as when intermodal interaction resolves a discrepancy that otherwise would lead to adaptation (Welch and Warren, 1980).

Crossmodal illusions are the most striking cases of multimodal interaction. As in the study of perception more generally, such illusions are evidence of the processes involved in normal functioning. They help to reveal perceptual principles and organization that otherwise are disguised. Indeed, explaining crossmodal illusions contributes to explaining multimodal unification and integration and draws attention to the most philosophically fruitful respects in which perception is multimodal. Interaction among the senses shows that perceiving involves not just the independent operation of discrete sense modalities. Understanding perception thus requires more than assembling accounts of the separate modalities considered in isolation from each other—it requires recognizing a unified multimodal perspective.

3 Crossmodal illusions

What you perceive with one modality ultimately can impact what you perceive with another. This is uncontroversial. If you see an object, you might now reach out to touch it and later taste it. When you hear a sound, you might turn and see its source. What is surprising is how directly one sense can impact another. For instance, just as a cue presented on one side of your visual field can, by attracting selective attention, enhance response time and accuracy for visual targets presented on that side (Posner, 1988), an *auditory* cue presented to one side enhances response time and accuracy for *visual* targets presented on that side by attracting selective attention *across modalities*. Crossmodal cuing also operates in reverse and across other pairs of modalities (see Spence et al. 2004, for review).

More strikingly, stimulation in one sense modality can cause an illusory experience in another. Vision, for instance, affects *spatial* aspects of experience in other modalities and frequently leads to illusions. First, it commonly causes illusory experiences of spatial location in other senses. Ventriloquism is an illusory experience of spatial location in audition that is caused by the visible location of an apparent sound source. Though best known from the case of hearing a puppeteer "throw" a voice, the ventriloquist illusion does not require speech. A flash can impact where you hear a concurrent beep. Ventriloquism is an illusory auditory experience caused by something visible (see Bertelson, 1999). (Many are surprised to learn that ventriloquism does not involve throwing sound; it is an auditory perceptual illusion as of sound coming from a place where its apparent source is seen.) Vision also captures proprioceptive location. For example, seeing a displaced image of your hand, or seeing a rubber hand, illusorily shifts where you *feel* your hand to be (Hay et al., 1965; Pick et al., 1969). Vision, however, also causes spatial illusions of shape, size, and orientation in experiences associated with other modalities (see, e.g., Rock and Victor, 1964). In general with such crossmodal effects, differences across subjects exist for a given effect, but individual subjects are quite consistent across trials. Individuals even are consistent in their susceptibility, and in the magnitude of the effect, across different kinds of cases.

Each of the cases above is compatible with the thesis that *vision wins* or dominates another modality whenever a conflict exists. Deference to vision reinforces the impression that it is the primary perceptual modality, and perhaps vindicates visuocentric thinking about perception in philosophy and cognitive science.

Crossmodal effects on perceptual experience, however, are not limited to vision's impact upon other modalities, and they do not concern just spatial features. Audition, for example, leads to visual recalibrations for temporal properties. The beginning of a sound can alter when a light seems to switch on so that the light's onset seems synchronous with the sound's. A sound's duration can alter the perceived duration of a visual stimulus, and a quick beep can make a moving visible target appear to freeze (Vroomen and de Gelder 2000). Sound also can alter visibly perceived rate and even temporal order.

One might think that crossmodal effects are limited in the following way. Vision impacts other modalities concerning spatial features, and audition impacts other modalities for temporal features. But, other modalities, including touch, can impact the experience of temporal properties in vision, and audition even interacts to some extent with proprioception for spatial features (proprioception tends to win, but not always).

Moreover, other kinds of perceptible features drive crossmodal interactions and illusions. Qualitative characteristics can generate illusions across modalities, as when audition alters tactile experience of texture (Jousmäki and Hari, 1998; Guest et al., 2002) or a smell alters taste. The McGurk effect is a more striking example that involves speech perception (McGurk and Macdonald 1976). When presented with the sound of the bilabial /ba/ phoneme (pronounced with the lips together) along with incongruent video of a speaker articulating the velar /ga/ phoneme (pronounced with the tongue at the back of the palette), many listeners report experiencing clearly the sound of the alveolar /da/ phoneme (pronounced with the tongue near the front of the palette), a kind of average or compromise. The effect quickly stops when you look away from the mouth.

One modality also can influence causal impressions in another. Two disks that travel toward and past each other can look either to *stream* past one another or, sometimes, to *bounce* and rebound from one another. A sound played when the disks coincide helps resolve the ambiguity and leads to a far higher portion of bounce percepts (Sekuler, et al., 1997).

Finally, a fascinating crossmodal illusion recently discovered by Shams et al. (2000, 2002) involves audition's impact upon vision even when stimulation to each is unambiguous and even though there is no *obvious* conflict. In the sound-induced flash illusion, a single flash accompanied by two brief beeps causes many subjects to experience not just two beeps but also *two* flashes. A sound impacts the number of events you visually experience. The effect continues for three and sometimes four beeps. But, a visual stimulus is required: the sound alone does not generate a visual experience. The illusion is asymmetric: vision does not impact audition in this way. Shams et al. report that this is a visual perceptual illusion caused by a sound, and that their results cannot be explained by extra-perceptual cognition or by subjects' adopting a strategy to help respond to ambiguous or conflicting experiences.

4 Explaining crossmodal illusions

4.1 Multimodal organizing principles

Crossmodal illusions take place when stimulation of one sense modality impacts perceptual experience associated with another. Sometimes, then, a process connected with one sensory system impacts a process connected with another. That means information from one sense can change how another responds.

The sound-induced flash illusion and a battery of other cases show that the predisposition towards intermodal perceptual recalibration and reconciliation is quite strong. That crossmodal effects are so prevalent indicates that crossmodal biases and illusions are neither aberrations nor mere quirks. This suggests that, if they do not just stem from accidents or miscueing, crossmodal illusions do result from processes governed by *intermodal* perceptual organizing principles. It also suggests that such intermodal organizing principles apply quite generally. For one, they constrain perception even under ordinary conditions when ambiguity and conflict are absent from sensory information and no recalibration occurs. For another, they apply across different sets of sense modalities. Perhaps the principles governing interactions among modalities even all share some general function or rationale. Crossmodal effects in that case reveal that inescapably multimodal processes are pervasive in perception.

It is helpful to contrast these multimodal perceptual effects with *synaesthesia*, another case in which stimulation in one sensory system impacts experience ordinarily associated with another. In synaesthesia, subjects enjoy, to mention a few examples, color experiences in response to sounds, sound experiences in response to colors, texture experiences in response to tastes, or tactile experiences in response to sounds (see, e.g., Baron-Cohen and Harrison 1997; Cytowic, 2002). In contrast to crossmodal perceptual illusions, synaesthesia is relatively rare. Synaesthesia that involves qualitative perceptual phenomenology occurs in roughly 1 in 2000 persons. More to the point, synaesthesia is a kind of experiential quirk that results from highly contingent facts about a person's sensory wiring, such as the proximity of functionally distinct pathways or brain regions. Synaesthesia differs in another important respect from crossmodal illusions. The processes responsible for synaesthesia, in contrast to those responsible for crossmodal illusions, always lead to illusions. Synaesthetes do not literally hear colors or taste roughness. There is no connection between the colors of things and the colors synaesthetes experience as a result of hearing sounds. The items synaesthetes perceive usually just lack the qualities synaesthetically experienced. How, then, could synaesthesia involve *perceptual* principles? In contrast, crossmodal illusions result from intermodal biases and recalibrations that are not accidental. First, they concern features that actually are present and available to more than one modality. They impact the experience of features we perceive, while synaesthesia involves outright property hallucinations. Second, they in fact help to improve the accuracy of our perceptual responses given information from multiple sensory sources. Not only do hands commonly have the locations we both see and proprioceptively feel them to have, but vision's impact on other modalities concerning spatial features resolves ambiguities and corrects or minimizes perceptual conflicts and errors. Such ambiguities, conflicts, and errors might stem from differences to the resolving power, accuracy, or noisiness of sensory stimulation. That audition overrides vision for certain features or under certain conditions enhances our capacity to perceive temporal characteristics. Intermodal processes commonly even help to avoid illusion, as when vision corrects a front-back confusion in auditory localization. So, although they sometimes lead to illusion, the principles that govern intermodal processes nevertheless are intelligible as advantageous. Synaesthesia, in contrast, does not stem from processes that in general enhance the capacity to perceive. Synaesthesia, unlike ordinary intermodal biases and recalibrations, is not in this way intelligible as adaptive.

4.2 Crossmodal triggers and mechanisms

What, then, are the *mechanisms* by which intermodal bias is exercised and recalibration takes place, and what are the *principles* that govern multimodal interaction and conflict resolution? Do any *general* principles exist that govern all crossmodal effects (see, e.g., Handel, 2006)? Detailed answers to these questions currently are scarce. However, some philosophically relevant patterns are emerging.

Three primary factors are known to trigger or influence multimodal interaction (for review, details, and discussion, see, e.g., Welch and Warren, 1980; Stein and Meredith, 1993; Spence and Driver, 2004; Calvert et al., 2004; Handel, 2006). *Spatial* and *temporal* information is most important. Being in the same place at the same time is the strongest trigger to intermodal coordination. Spatiotemporal proximity commonly cues recalibration when different senses disagree, whether the disagreement concerns space, time, or some other attribute. Perhaps this stems from two more basic principles: that two distinct things cannot occupy just the same place at the same time (cf. Bedford, 2004), and that a single thing cannot have conflicting features. Crossmodal interactions and illusions weaken predictably once spatial and temporal information from different senses begins to diverge. Increasing disagreement beyond some threshold leads to perceptual separation with biasing toward the dominant modality, followed by complete dissociation without biasing. Spatiotemporal information figures among bottom-up influences on multimodal processes. One interesting upshot of the spatial parameters on crossmodal interactions concerns the representation of space in different modalities. Austen Clark (forthcoming) argues that to explain crossmodal cuing of spatial selective attention, a common spatial framework must be represented as such across modalities. Perhaps a similar argument could be constructed for temporal features.

Both active and passive *attention* also impact the strength of crossmodal interactions and illusions. Whether directed intentionally or not, attention enhances multimodal processing and strengthens the inclination to reconcile discrepant information from distinct sensory sources. Attention facilitates feature binding, so perhaps some multimodal processes depend upon feature binding. Attention counts among top-down perceptual influences on multimodal processes.

Finally, *compellingness* might play a role in multimodal perception. For instance, hearing the sound of a teakettle while seeing spatially offset steam causes a stronger crossmodal spatial illusion than hearing the sound of a bell while seeing a spatially offset bell that does not vibrate in the expected way (Jackson, 1953; cited in Handel, 2006, p. 407). Top-down cognitive factors might therefore modulate the impact of more basic features, such as spatial and temporal characteristics, on crossmodal biases and illusions. These cases also show that we need to be careful to distinguish the effects of automatic perceptual processes from those of response strategies adopted to deal with ecologically suspicious experimental tasks (see de Gelder and Bertelson, 2003).

Physiologically, perhaps surprisingly, the mechanisms that drive crossmodal effects include connections between processes formerly thought to be modality-specific. Areas of the brain previously believed to be dedicated entirely to unimodal tasks now are known to be activated by extramodal stimuli or to respond to multimodal input. Such effects can occur quite early. For instance, V1 in the early visual system receives auditory information during the sound-induced flash experiment (Watkins et al., 2006). Furthermore, there are neurons dedicated to receiving multimodal input, while certain areas of the brain, such as superior colliculus, play a critical role in consolidating multimodal information (see Stein and Meredith, 1993). Top-down influences, such as processes associated with selective attention activated by one modality, also impact multiple modalities (see, e.g., Driver and Spence, 2000). Multimodal interactions thus occur in various ways at different stages in the perceptual process, notably including those formerly thought unimodal (the essays in Calvert et al., 2004, provide a comprehensive review).

4.3 Why multimodal interaction?

Given these general triggers, what governs or determines how a multimodal interaction, recalibration, or illusion unfolds? For instance, why does vision sometimes win, while audition at other times impacts visual experience? What determines the strength of an illusion? Among the candidates for explaining particular patterns of bias, recalibration, and illusion are considerations such as which modality is more appropriate for the kind of information or feature in question, which modality is more accurate, and what perceptual outcome enhances the likelihood of correctness. Consider them in turn.

According to the *modality appropriateness hypothesis*, the modality that is most *inherently appropriate* for a particular kind of feature biases or dominates another when disagreement exists over that feature (see, e.g., Welch and Warren, 1980). For instance, vision dominates audition when disagreement exists about space, and audition wins when the conflict concerns temporal features. Inherent appropriateness might appeal to physiological characteristics that make vision capable of fine spatial resolution and audition capable of fine temporal resolution. Sometimes, however, the modality that intuitively is not most appropriate for a given feature biases the one that is. Sound can impact felt texture, and vision biases speech perception, as in the McGurk effect. Furthermore, biasing effects can be reversed. Blurring vision or adding visual distractors increases auditory dominance for spatial features by decreasing visual reliability. Altering the reliability of a modality for a given feature changes dominance patterns. If appropriateness depends upon reliability, appropriateness is not an inherent characteristic of a sense modality.

In light of this, perhaps *deference to the most reliable modality* governs crossmodal recalibrations. If so, audition wins when it most reliably indicates a given feature, and vision wins when it does, independent from which is most inherently appropriate for a given feature. This is not quite right, either, since crossmodal recalibrations come in degrees. In fact, four patterns are common when information about some feature, such as spatial location or temporal onset, disagrees. Consider, for example, visual and auditory information about spatial location. First, when the information differs wildly, we simply see one thing in one location and hear another thing in another location, with no attempt to resolve the disagreement. Second, when information differs significantly but not wildly, we might see something in one location and hear another thing in another location. But, in this case, one or both locations might be shifted or biased to some extent towards the other. Third, when the difference is less drastic, we might see and hear something in a common spatial location that is an average or compromise of the information stemming from the two modalities. Finally, we might see and hear something in a common spatial location that coincides with the actual location of the stimulus to one modality or the other. The first pattern involves no crossmodal bias, while the second involves bias without convergence. The third pattern involves convergence with mutual bias, while the fourth involves complete deference or dominance. The lesson is that total dominance in favor of one modality or the other in situations of conflict is not the rule. Ventriloquism, visual capture of proprioception, and McGurk effects each involve some measure of compromise.

A promising reply is that crossmodal biases and recalibrations *enhance* the reliability of perception in the face of noisy, fallible, or imprecise sensory stimulation. What does this involve? In ordinary cases when different senses agree, experience in each modality simply becomes more salient and detection improves. But when information conflicts and the senses do not agree, complete resolution in favor of one modality is a limit case. Frequently, averaging or weighting or some form of compromise occurs. To explain this, distinguish the reliability of a modality for a given kind of feature from the reliability of perception overall for that feature. Deferring completely to the modality that is most reliable for a given feature does not always lead to the optimal or most reliable perceptual result. Rather, the likelihood of some worldly situation in the face of the given visual and auditory stimulation, for example, is determined by the reliability of each modality relative to the other for that feature. Enhancing the reliability of the perceptual results thus should involve a weighting of evidence from the different sources. In fact, intermodal biasing and recalibration frequently conforms to a *weighting function* that incorporates the relative reliability of multiple modalities in order to determine the strength of the contribution from a given modality to the perceptual compromise. Thus, for instance, visual cues might to a great extent impact the auditory experience of spatial location, while auditory cues make a significantly smaller contribution than visual cues to the visual experience of spatial location. When the reliability of each modality matches, the perceptual result in each involves averaging. Weighting functions should predict the degree to which biasing takes place, and they should explain perceptual errors caused by multimodal recalibration. In addition to the spatial and temporal crossmodal illusions, for instance, weighing both visual and auditory contributions nicely captures why a sound biases an ambiguous visual display from mostly streaming to mostly bouncing percepts, or why visual cues cause the auditory McGurk illusion even though the auditory stimulus is unequivocal. In the latter case, deference to the auditory cue arguably would make more sense given audition's strength at resolving spectral and temporal information about speech sounds. However, the compromise that occurs when a visual cue accompanies the auditory signal enhances the reliability of the resulting percept, given that each modality bears information about spoken phonemes. Similar explanations have been applied to other crossmodal illusions, including the sound-induced flash illusion (see Shams et al., 2005). Weighting functions that enhance perceptual reliability thus show promise in explaining multimodal adjustments across a range of cases (see Handel, 2006, ch. 9). A critical but open question is just how closely crossmodal recalibrations conform to what is statistically optimal.

5 Conflict, content, and phenomenology

Commonly, sensory stimulation and processes associated with one perceptual modality alter processes associated with another modality. This results from perceptual principles that govern even ordinary cases in which no illusions occur: multimodal triggering cues might be absent; or, the triggers might be present, but, if no conflict exists, there is no need for reconciliation; or, multimodal interactions might correct sensory noise or inaccuracies that would have caused perceptual error. Taking advantage of several modalities enhances the reliability of perception.

Explaining multimodal effects and principles impacts a number of other topics in the philosophy and cognitive science of perception. I want to shift from explaining multimodal interactions to explaining their impact on how we understand perception. The rest of this section argues that multimodality has consequences concerning perceptual content and phenomenology.

5.1 Why conflict matters

Different senses sometimes bear conflicting information about some feature. Consider what happens in cases of intermodal conflict. The telling fact is that divergent information from different senses so frequently is *reconciled*—conflicts are resolved. It is striking that biasing, reconciliation, and convergence take place at all when different senses bear different information. Why not simply settle for different perceptions with different senses?

Explaining patterns of multimodal interaction that alter and reconcile information from different senses as conflict resolution requires recognizing that information from different senses, which may be encoded in different sense-specific ways, is information about something common. Resolving crossmodal conflicts requires that perceptual systems in effect treat information from different senses as information *about the same thing*. Since the conflicting information is treated as information about something available to more than one sense modality, multimodal reconciliation exhibits a perceptual concern for common objects across the senses.

This has been reflected in what researchers have called *unity assumptions* that govern crossmodal interactions. Unity assumptions embody the criteria according to which perceptual systems determine whether to treat stimulation from distinct senses as belonging together, where belonging together means being a possible subject for multimodal recalibrations.

The importance of this point has been underestimated. Multimodal interaction is not just random or accidental, since it resolves or minimizes differences among modalities. It reconciles information from different senses by recalibrating responses from different sensory systems in a rule-governed way. Reconciliation, however, requires a conflict in divergent information, and genuine conflict requires disagreement. Disagreement requires a common subject matter, or it is merely apparent. Even apparent disagreement requires the presupposition of a common subject matter. So, to explain what perceptual systems do in multimodal contexts as recalibrating in order to reconcile conflicting stimuli or divergent information from different senses requires that perceptual systems treat information drawn from different senses as concerning something common—as information about the same items or features.

Understanding crossmodal illusions and intermodal interactions sheds light upon ordinary perception because similar considerations help to explain multimodal integration even when no recalibrations or illusions occur. Conflict resolution requires conflict detection, which requires determining when information from different sources is commensurable. Since they require conflict detection, the mechanisms that trigger intermodal recalibrations depend upon mechanisms to determine whether perceptual information from different senses is commensurable. But commensurability determines what ought to be unified or integrated. The need to unify or integrate, after all, is what makes conflict a problem that needs resolving. If the multimodal processes responsible for determining when information from different sources is commensurable are among those that determine the content and phenomenology of perceptual experiences, then they provide a glimpse into part of what is responsible for the perceptual sense that experiences associated with several modalities belong together or are unified in that they perceptually seem to concern common individuals and features.

5.2 Common content

Reconciling information from different senses demonstrates a perceptual concern for items or features that are common to different modalities. It is not enough that information from different senses happens to share a subject matter. Even though distinct senses bear information that differs in format and content, they must be treated as bearing *commensurable* information to drive reconciliation. Remarkably, amid so much divergent information across the senses, certain information streams are treated as possibly conflicting or agreeing. Conflict resolution thus implements a grasp upon information from different senses *as* information about the same thing. The multimodal activity patterns involved in recalibrating and reconciling information drawn from different senses amount to the exercise of a kind of perceptual grasp upon sources of sensory stimulation that are common across modalities.

Such a perceptual grasp cannot be entirely unique or specific to a given modality, since it requires, for instance, treating auditory information as sharing a source or subject matter with visual information. So, this grasp cannot be characterized entirely in modality-specific terms. Rather, given that multimodal principles are deployed in response to information drawn from several modalities, and given that they involve treating sensory information as sharing a common subject matter, multimodal processes involve either a multi-modal or modality-independent (amodal) *way* of grasping the common sources of sensory stimulation. Crossmodal illusions are evidence of this grasp.

If multimodal processes exercise or implement a perceptual grasp upon common sources of sensory stimulation that is not modality-specific, this supports the claim that there exists a variety of perceptual content that cannot be captured in modality-specific terms. Some theories of content require a causal connection. Representations that result from multimodal recalibrations are causally connected to things or features accessible to more than one sense, and in a way that runs through more than one sense. Some theories require reliable or counterfactual supporting connections. Multimodal processes reliably identify and track items and features accessible to multiple modalities—auditory location, for instance, frequently responds accurately to changes in visual information about location. And they do so in a way that is transferable across modalities—you do not lose track of a basketball when you simultaneously grab it and stop looking at it. All theories of content entail satisfaction or veridicality conditions. Multimodal processes, including biasing and recalibration, ground perceptions with correctness conditions that cannot be characterized exhaustively in terms that are specific to a given modality. For instance, perception that involves multimodal recalibration is veridical only if a single feature or individual, which cannot intelligibly be ascribed conflicting characteristics, is responsible for the stimulation to distinct senses. It is not enough in such cases that an auditory-location and a visual-location, or that a visual-object and a tactual-object, are perceptually represented. The auditory-location and visual-location must be identified, as must the visual-object and tactual-object. Thus, many crossmodal illusions are not just misperceptions in a given modality of some feature; they also mistakenly *identify* what is perceived through different modalities. For instance, visual capture of the felt location of your hand is not just a proprioceptive illusion of spatial location; it also involves mistakenly identifying what you see with what you feel. Ventriloquism is not just an auditory spatial illusion; it is perceptually misidentifying the seen puppet as the sound source. Ventriloquism would involve an illusion even if it literally involved "throwing" a sound to the puppet's location.

What sorts of things might figure among such contents? The features and items in question likely belong to two groups. First, broad structural features such as space and time serve as basic triggers that drive intermodal interactions. They frequently are the respects in which senses disagree, so spatial and temporal features must be commensurable across modalities. This suggests that spatial and temporal features figure in perceptual contents common across modalities. Spatial and temporal information alone is not enough, however. That vision represents one place and audition represents another does not explain reconciliation. Spatial and temporal information is information about something. Explaining reconciliation requires recognizing that the spatial or temporal information ascribed *to* something conflicts. Spatial and temporal conflict requires items or individuals to which spatial and temporal features are attributed. What are the candidates for items that might figure in contents ascribed to different modalities? Individuals, such as objects and events, play a critical role in multimodal perception. They may have spatial or temporal locations, and features such as shape, rate, or duration. They also may have special features like color, taste, or loudness that are accessible to a unique modality. They are good candidates among the multimodal targets of attention since, plausibly, they are tracked across modalities, as when you turn to see the source of a sound. These individuals thus, plausibly, also anchor perceptual demonstratives and are available as common subjects of empirical thought. If explaining crossmodal recalibration requires positing a perceptual grasp upon common items about which different senses disagree, individuals accessible to multiple modalities are a component of perceptual content that cannot exhaustively be characterized in modality-specific terms.

5.3 Shared phenomenology

I have been discussing perceptual processing, explanation, and content. What about perceptual experience? Some *intentionalist* theories hold that perceptual phenomenology supervenes upon perceptual content (see, e.g., Harman, 1990; Dretske, 1994; Tye, 2000; Byrne, 2001). If multimodal perception shows that perceptual experiences associated with different senses can share content, intentionalism therefore implies that they also must share perceptual phenomenology. This consequence has been used to argue against intentionalism (see, e.g, Lopes, 2000; O'Dea, 2006). It is intuitive to think that we perceptually experience things and features in ways that are *distinctive* to each modality. Even if we can see and touch the same thing, and even if something common must figure in the contents of sight and touch, it is natural to suspect there is a distinctively visual way of experiencing it and a different, recognizably tactual way of experiencing it. That is, even if common sensibles figure in common contents, there are still different ways of perceptually experiencing or entertaining them with different senses. Some respond with an intramodal restriction to the intentionalist thesis. Despite intuition, however, multimodal cases ground a different intentionalist response.

Consciously perceiving something to be blue, or to be illuminated on its left side, or to occlude part of a surface, surely involves a distinctively visual experience. Perceiving something as high-pitched, or as loud, involves a distinctively auditory experience. Since, at a time, in everyday conditions, one's visual or auditory experiences arguably all involve at least some modality-specific phenomenological features, the overall phenomenology of visual or auditory experience arguably is distinctive. Nevertheless, it is mistaken to hold that perceptual phenomenology is exhausted by modality-specific phenomenology. For ordinary perceivers, in whom perceptual principles govern multimodal bias, interaction, and reconciliation, perceptual experiences include phenomenological characteristics that must be shared by or common to experiences associated with different senses.

So far, I have argued that we perceive common items that are perceptually grasped or represented as common across multiple modalities. This is required to explain conflict and reconciliation. But, one might argue that common content is compatible with distinctive phenomenological character for every pair of perceptual experiences that occur in distinct modalities. Crossmodal illusions then might stem from lower-level causal and informational processes that adjust phenomenologically distinctive experiences in one or more modalities. That causal interaction takes place between processes that lead to experiences in different senses, and that explaining those interactions requires positing grasping or representing as of common items, might not be reflected as such in phenomenology. Causal influence does not imply a constitutively dependent experience. On a traditional picture, each sense contributes something entirely distinctive, and sense-specific contributions exhaust overall experience. Perhaps the traditional picture is safe.

However, evidence exists that multimodal sharing is reflected even in perceptual experience and that perceptual phenomenology is not exhausted by modality-specific characteristics. Consider cases in which *intermodal* binding takes place. Just as distinct features drawn from a single modality, such as color and shape, are *intramodally* bound in visual experience—a single perceptible thing seems to have both color and shape, and to be distinct from other things that have color and shape—features drawn from different senses, such as visible shape and felt texture, sometimes perceptibly belong to a single item thanks to intermodal binding (see, e.g., Pourtois et al., 2000). The very same thing perceptually seems to have features drawn from more than one sense modality. That which is experienced to have a visible feature is experienced also to have a tactile feature (or an audible feature). That which bears the visible feature is identified with that which bears the tactile feature, so a single thing is experienced as having both visible and tactile features. In order to capture the phenomenology of cases in which intermodal binding takes place, some aspect of perceptual phenomenology must account for the experiential sense in which the item seen is the very same item touched. That is, some aspect of perceptual phenomenology is common to the visual experience and the tactual experience. It corresponds to whatever is experienced as having both visible and tactile features. This aspect of perceptual phenomenology might itself be multimodal in nature, or it might have an entirely amodal character. Either way, distinctive sense-specific characteristics do not jointly exhaust the phenomenology of perceptual experience. Either experiences associated with different senses share at least some aspects of phenomenological character, or perceptual experiences are not exhausted by distinctive phenomenological characteristics that are specific to individual modalities (see O'Callaghan, 2008b).

This, however, implies that considering purely visual cases and invoking properly visual characteristics does not suffice to characterize the phenomenology even of visual perceptual awareness. That approach abstracts from the context of other modalities and thus ignores their contributions to integrated multimodal perceptual experience. Not only do we assess vision and visual experience from the perspective of multimodal awareness, but visual experience itself cannot exhaustively be characterized, understood, or explained without comprehending its role in integrated multimodal perceptual experience. A complete account of vision and visual experience requires constitutive relationships among modalities.

6 The senses

Multimodal perception, especially crossmodal interactions and illusions, bears on how we understand the nature of the senses.

6.1 Are the senses modular?

One might be tempted to think the senses are *modular* (for instance, in the sense of Fodor, 1983, 2000). The senses often are mentioned as examples of mental modules. Suppose you hold that perception in general is modular and thus impenetrable by beliefs and other forms of cognition. And suppose you hold that the processes associated with each sense are specific or dedicated to that sensory domain and that they operate mostly independently from other senses. Then you might hold that audition, for instance, is informationally encapsulated and impenetrable by information from other sensory domains. You might count audition among the modules.

This view is far too strong. Multimodal processes involve causal interaction and information transmission among sensory systems. Perceptual processing in vision is shaped and changed by auditory information. Connections between sensory systems exist at numerous physiological locations and at different functional or computational stages. Shimojo and Shams (2001) argue explicitly against what they call the historically prevalent view that the sense modalities are distinct modules on the grounds of multimodal plasticity and interaction.

However, it is difficult to find work that advocates such a strong version of the modularity thesis for the senses. Some hold that only an early part of the visual system is modular (e.g., Pylyshyn, 1999). Some hold that even early vision's modularity depends upon a number of distinct systems working in concert that detect, for example, color, motion, edges, and form. In that case, early vision itself employs a number of modules. Perhaps it is no surprise that complete sensory systems fail to act as strict modules. Multimodal effects make trouble only if you believe that all perceptual modules are modality-specific. But on a weakened modularity view, why maintain this? Perhaps, while vision, audition, smell, taste, touch, and proprioception are not strictly modular, functional modules responsive to common multimodal or amodal features and individuals, such as space, time, objects, and events (and perhaps even perceived language), might exist. This would seem necessary to uphold even a less-than-massively modular view of the mind on which perceptual systems are distinguished from higher cognition by their modularity (see, e.g., Fodor, 2000). The important upshot is that modules specific to vision, audition, and the rest do not exhaust perceptual capacities. Certain critical perceptual processes require input from several sensory systems and impact experience associated with multiple modalities. In addition to any sense-specific functions, perception involves multisensory tasks.

6.2 Individuating the senses

Does the multimodality of perception challenge even the common understanding that we have five (or so) different senses? I have assumed that we can individuate the senses in a way that corresponds to folk notions of seeing, hearing, feeling, tasting, and smelling. But multimodal perception casts doubt on several competing accounts of the senses.

Four kinds of proposals for individuating the senses have been advanced. Respectively, they individuate the senses in terms of their *objects*; in terms of characteristics of *experience*; in terms of *physics*, *physiology*, or *function*; and by mere *convention*. How do these accounts fare given the results discussed in this chapter?

Can we individuate the senses in terms of their intentional *objects* (or *sensibles*)—what is sensed or perceived thanks to that modality (a view commonly attributed to Aristotle; cf. Grice, 1962)? Since some features and individuals, such as things we both see and hear, figure among the objects of more than one modality, citing objects alone does not suffice to distinguish the senses. This forces us to appeal to proper sensibles or objects—things perceived only with one sense. But knowing whether something is available to one sense or two senses assumes we can count the senses. Furthermore, each sense has several proper sensibles, such as pitch and loudness for hearing, or hue and brightness for vision. Thus, in light of multimodal considerations, individuating senses in terms of their objects results either in violations of common sense distinctions, such as seeing and hearing, or in too many senses.

Perhaps, in light of the problem of common sensibles, we can appeal to distinctive characteristics of *experiences* associated with each sense to individuate the modalities, as Grice (1962) suggests. If the phenomenological conclusion of section 5.3 is correct, then this account faces similar problems. If there are aspects of phenomenology common among modalities, such as seeing and hearing, then neither the experience of seeing nor of hearing is exhausted by entirely distinctive phenomenological characteristics. Therefore, no phenomenological characteristic is shared by all and only visual experiences. Thus, we cannot individuate the senses experientially or phenomenologically unless we acknowledge some additional perceptual modality beyond the five traditional senses. This additional modality then could accommodate a distinctive experience of, for instance, intermodal binding, while distinctive modality-specific experiences remain for vision, audition, and the rest. This, however, does damage to the claim that colors and shapes, for example, visually seem to belong to things we see rather than to things perceived with some other extra-visual modality, and that textures and form tactually seem to belong to things we feel. Some also might find it odd to say that an amodal or multimodal *modality* exists in addition to the traditional senses.

Heil (1983) suggests we individuate the senses in terms of the physical energy to which they respond; Keeley (2002) suggests we individuate the senses in terms of organs evolutionarily dedicated to picking up information of a certain kind. Neither of these accounts strictly conflicts with the multimodal results discussed above, but once we take them to apply to perceptual systems, multimodal processes become a problem. The end sense organs are not hooked up to functionally discrete systems that are dedicated to responding to unique information or features in the way that the eye responds to light, the ears to mechanical pressure waves, and the nose to chemicals. Multimodal perceptual processes discern and respond to constancies and commonalities in stimulation across sense organs. Individuating senses either by end organs or by physical energy types therefore does not suffice to individuate corresponding perceptual systems. Since the senses on such accounts are not *perceptual* modalities, their theoretical interest to psychology and philosophy is minimized. Seeing interests philosophers and cognitive scientists not just because it involves having an organ that responds to light, but because seeing is a way of perceiving.

Matthew Nudds (2003) argues that, given trouble individuating the senses, we should say that the senses are conventional categories we treat as different ways of perceiving. Nevertheless, knowing that someone is seeing is informative because it suggests to us the kinds of things they are likely to perceive, such as colors, shapes, and objects. Knowing someone is hearing makes it more likely that they perceive pitches, sounds, and events. But we should resist thinking the senses correspond to physiologically or psychologically real kinds. The senses, on this view, are no more than folk psychological concepts with limited value to empirical science. Though unsettling, and though Nudds's argument from the failure of extant accounts does not rule out deeper similarities and differences that ground conventional distinctions among the senses, multimodal aspects of perception might lend support to this line of thought.

In light of this, perhaps we should restrict talk of the senses to low-level systems such as the eye and very early visual processes, which can be individuated as Heil and Keeley suggest. But these senses are not *perceptual* modalities, and they do not correspond neatly to categories of experience. Seeing, hearing, and the rest are ways of perceiving whose individuation might be messy in the way Nudds suggests. Perceiving, therefore, might be essentially multimodal in that it essentially involves principles and processes that span multiple sensory systems.

7 Multimodality and perception

In summary, crossmodal illusions demonstrate that perception involves interactions among processes associated with different modalities. Patterns of crossmodal bias and recalibration reveal the organization of multimodal perceptual processes. Multimodal interactions obey intelligible principles; they resolve conflicts; they enhance the reliability of perception; they are not mere quirks or accidents. Multimodal processes also demonstrate a concern across the senses for common features and individuals, for the following reason. The intermodal biasing and recalibration responsible for crossmodal illusions requires that information from sensory stimulation associated with different senses is grasped or represented as being commensurable. Since conflict resolution requires a common subject matter, commensurable information from different senses shares or traces to a common source. Crossmodal processes thus amount to the exercise of a principled perceptual grasp upon the common sources of sensory stimulation across modalities.

Further philosophical work is needed to characterize the varieties of multimodality and their bearing on traditional theories of perception, perceptual experience, and the senses. The implications for research methodology are clear: studying the various modalities in isolation reveals just the surface of the story about perception. Philosophical and empirical work thus should not proceed on the assumption that it is possible to understand perception and perceptual experience in terms of a single modality entirely in isolation from the others. For one, it is not possible simply to extrapolate or translate claims about vision into claims about other modalities. Moreover, vision itself may resist an exhaustive understanding that does not appeal to non-visual modalities. Without recognizing perception's deep multimodality, we overlook the role of the several senses in perception.

What does this mean for how we understand the nature of perception? Handel (2006) says perceiving is about solving *correspondence problems* through the detection of contrast and change. A correspondence problem is one of identifying an individual or feature, either at a time or over time, given sensory information that varies from location to location and from moment to moment. (The original correspondence problem concerns how to reconcile two retinal images to yield information about depth.) Perceiving, on this view, is about using contrasts and changes in noisy and variable sensory stimulation to carve up a sensible scene into individuals that bear relatively stable characteristics.

One important lesson of multimodal effects is that an analog of the correspondence problem within a modality holds between modalities. Perceiving involves determining that what you hear is what you see, that the object you feel to be in your hand is the one you see, or that what tastes bitter is what you feel on your tongue. An empirical account of perception should explain how this is accomplished. Philosophers and cognitive scientists should ask what makes the accomplishment significant.

Here is a start. One upshot deals with the relationships among experiences commonly associated with different modalities. Just as perceiving colors and visual-objects involves detecting constancies in hue and shape despite variations and changes in appearance due to lighting and viewing angle, perceiving ordinary objects and events implicates constancies in features detected with different senses, across changes to the modality with which one perceives. Without appeal to such constancies, shifting sense modalities would perceptually seem to result in altogether different objects of experience. If different modalities shared no intentional objects or features, perceptual experience would in one noteworthy way seem fragmented. We would perceive no relationships among things experienced through different modalities. Our sense of the cohesiveness of the world as we perceive it through different senses therefore depends upon our perceptually keeping track of common items and features across different modalities. One very critical aspect of perceptual experience thus stems from multimodal functioning. Objectual unity is tied to our capacity to detect constancies and solve correspondence problems across modalities.

Another upshot deals with the nature of experiences associated with a given modality. We perceptually identify and keep track of things despite contrasts in sensory information and in presentation across modalities. This, I argued, grounds varieties of perceptual content and phenomenology that are common to perceptual experiences associated with different modalities. It follows that characterizing perceptual content and perceptual phenomenology requires appeal to terms beyond those that are proprietary to a given modality. Characterizing an ordinary occasion of perceiving by means of audition therefore requires appeal to terms that are not inherently auditory. Talk of sounds, pitch, timbre, loudness, and audible-location, whose perceptual significance is purely auditory, needs to be supplemented with talk of sound sources—objects that make sounds, or events of sound production—and locations shared with vision. A similar claim holds for vision. We might therefore be justified in saying that we can visually perceive something *as* the sort of thing that could have auditory, tactual, olfactory, or gustatory qualities, rather than as something with visual significance alone. Likewise, one might hear something *as* the sort of thing that could be seen or touched. So, the experience of perceiving with one modality can embody *perceptual* expectations that implicate another. For instance, seeing a convincing hologram invites expectations that are violated when you learn you can put your hand through it without resistance. Such visual expectations are perceptual, rather than just results of extra-perceptual cognition or association. The scope of experiences associated with a given modality, as a result, might be far greater than traditional views imagine.

Solving crossmodal correspondence problems, on the approach I have discussed, requires a common amodal or multimodal code that is shared among modalities. The proprietary or distinctive aspects of sense perception and experience thus may distract from what is most noteworthy about perceiving. Perhaps it is in grasping multiple sensory perspectives as perspectives upon a common source that *the* world becomes available as a subject for attention, empirical belief, and action. If so, perceiving is a thoroughly multimodal affair.

Acknowledgments

References

Aristotle (1984). De Anima. In *The Complete Works of Aristotle: The Revised Oxford Translation*. Oxford University Press.

Baron-Cohen, S. and Harrison, J. E. (1997). *Synaesthesia: Classic and Contemporary Readings*. Blackwell.

Batty, C. (2007). *Lessons In Smelling: Essays on Olfactory Perception*. PhD thesis, MIT.

Bayne, T. J. and Chalmers, D. J. (2003). What is the unity of consciousness? In Cleeremans, A., editor, *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford University Press.

Bedford, F. L. (2004). Analysis of a constraint on perception, cognition, and development: one object, one place, one time. *Journal of Experimental Psychology: Human Perception and Performance*, 30.

Bertelson, P. (1999). Ventriloquism: a case of cross-modal perceptual grouping. In Aschersleben, G., Bachmann, T., and Müsseler, J., editors, *Cognitive Contributions to the Perception of Spatial and Temporal Events*. Elsevier.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Byrne, A. (2001). Intentionalism defended. *Philosophical Review*, 110.

Calvert, G., Spence, C., and Stein, B. E. (2004). *The Handbook of Multisensory Processes*. MIT Press.

Casati, R. and Dokic, J. (1994). *La Philosopie du Son*. Chambon.

Clark, Andy. (2006). That lonesome whistle: a puzzle for the sensorimotor model of perceptual experience. *Analysis*, 66.

Clark, Austen. (forthcoming). Cross-modal cuing and selective attention.

Cytowic, R. E. (2002). *Synesthesia: A Union of the Senses*. MIT Press.

de Gelder, B. and Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, 7.

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, 10.

Fodor, J. (1983). *The Modularity of Mind*. MIT Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. MIT Press.

Grice, H. P. (1962). Some remarks about the senses. In Butler, R. J., editor, *Analytical Philosophy, Series 1*. Blackwell.

Guest, S., Catmur, C., Lloyd, D., and Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, 146.

Handel, S. (1993). *Listening: An Introduction to the Perception of Auditory Events*. MIT Press.

Handel, S. (2006). *Perceptual Coherence*. Oxford University Press.

Harman, G. (1990). The intrinsic quality of experience. In Tomberlin, J., editor, *Philosophical Perspectives*, volume 4. Ridgeview.

Hay, J. C., Pick, H. L., and Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, 2.

Heil, J. (1983). *Perception and Cognition*. University of California Press.

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, 213.

Helmholtz, H. (1866/1925). *Treatise on Physiological Optics*, volume 3. Optical Society of America.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, 5.

Jousmäki, V. and Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology*, 8.

Keeley, B. L. (2002). Making sense of the senses: individuating modalities in humans and other animals. *The Journal of Philosophy*, 99.

Lopes, D. M. M. (2000). What is it like to see with your ears. *Philosophy and Phenomenological*

Research, 60.

Lycan, W. (2000). The slighting of smell. In Bhushan, N. and Rosenfeld, S., editors, *Of Minds and Molecules: New Philosophical Perspectives on Chemistry*. Oxford University Press.

Martin, M. (1992). Sight and touch. In Crane, T., editor, *The Contents of Experience*. Cambridge University Press.

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

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264.

Noë, A. (2004). *Action in Perception*. MIT Press.

Nudds, M. (2003). The significance of the senses. *Proceedings of the Aristotelian Society*, 104.

O'Callaghan, C. (2007). *Sounds: A Philosophical Theory*. Oxford University Press.

O'Callaghan, C. (2008a). Object perception: vision and audition. *Philosophy Compass*, 3.

O'Callaghan, C. (2008b). Seeing what you hear: crossmodal illusions and perception. *Philosophical Issues*, 18.

O'Dea, J. (2006). Representationalism, supervenience, and the cross-modal problem. *Philosophical Studies*, 130.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24.

Pick, H. L., Warren, D. H., and Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception and Psychophysics*, 6.

Posner, M. I. (1988). Structures and functions of selective attention. In Boll, T. and Bryant, B., editors, *Master Lectures in Clinical Neuropsychology*. American Psychological Association.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., and Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, 11.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22.

Rock, I. and Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science*, 143.

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature*, 408.

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14.

Shams, L., Ma, W. J., and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport*, 16.

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11.

Spence, C. and Driver, J., editors, (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press.

Spence, C., McDonald, J., and Driver, J. (2004). Exogenous spatial-cuing studies of human crossmodal attention and multisensory integration. In Spence, C. and Driver, J., editors, *Crossmodal Space and Crossmodal Attention*. Oxford University Press.

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

Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.

Tye, M. (2003). *Consciousness and Persons: Unity and Identity*. MIT Press

Vroomen, J. and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88.