

Cross-modal illusions and perceptual content

Abstract

Philosophical work on perception traditionally has focused upon vision. I argue that thinking about modalities other than vision furnishes challenging new puzzles, but fails to go far enough. Considering the relationships among perceptual modalities sheds light upon what is most striking about perception: its capacity to furnish awareness as of a world of things and events distinct from oneself. I begin with a puzzle from audition and argue that a class of recently discovered cross-modal illusions helps to resolve the puzzle. Such illusions, however, also reveal a dimension of content shared across perceptual modalities and tell against the traditional view according to which perceptual content consists in discrete, modality-specific contents.

Word count: 3486

1. Introduction

Philosophical thinking about perception has been driven primarily by attention to vision and to visual examples. Color, spectrum inversion, the waterfall illusion, blindsight, and change and inattention blindness are just a few examples in which vision has furnished the puzzles and delivered the intuitions relevant to assessing any philosophical theory of perception. Thinking about modalities other than vision bears fruit not only by challenging or confirming what we learn through thinking about vision, but also by adding new puzzles that shape our understanding of perception. Simply shifting to other modalities, however, does not go far enough toward abandoning the 'visuocentric' focus because it fails to reveal the most significant implications of considering multiple modalities in developing and evaluating theories of perception and perceptual content.

2. A puzzle about audition

I'll begin with a puzzle concerning sounds and audition. It is clear that, in a relatively innocuous sense, sounds are the immediate objects of auditory experience -- whatever else you hear, such as cars or crashes, you hear it in virtue of hearing a sound. Auditory experience is, nonetheless, *object-* and *event-*involving. You learn on the basis of

auditory experience that the glass has broken, that there is a bell in the room, or that the train is passing. One plausible view about how you learn this is that you hear the train, the bell, or the breaking of the glass. The experience seems to you to be an experience of a train, a bell, or a glass breaking. It is no accident that we speak about and classify sounds in just these terms. So, you hear a sound, and by or in hearing that sound, you hear the everyday object or event that is its source. This awareness may feel less ‘direct’ than your awareness of the sound or the awareness of the apple you enjoy on the basis of seeing its color and shape, but there is a sense in which it seems that one enjoys auditory awareness of a train, a bell, or a glass breaking in virtue of hearing its sound.

The puzzle is this: How could auditory experience, whose proper objects are sounds, which are distinct from ordinary objects and events, furnish perceptual awareness of things like trains, bells, and breakings?

The puzzle raises two closely related questions about the content of auditory perception. The first is, ‘How mediated is one’s awareness of ordinary objects and events in audition?’ That is, how remote is the sense in which one grasps such particulars as bells and whistles if that awareness is mediated by sounds. The second is, ‘How rich is the content of auditory experience?’ Simply, do those contents encompass bells and whistles, or merely sounds?

Defending an answer to these questions requires a story both about the contents of such awareness and about what grounds this particular form of mediated perceptual awareness. In short, we must explain how hearing could inform us perceptually about ordinary objects or events, which traditionally have not been considered among the objects of audition proper.

One might argue that the apparent perceptual awareness of ordinary objects and events is a mere illusion, and that the sound in fact mediates consciousness of objects and events only modulo some inferential or otherwise cognitive connection. But the phenomenology of audition seems for all the world to furnish experiential awareness of things and happenings beyond sounds. We reflexively act to orient toward or to avoid the source of a sound. Nonetheless, perhaps the puzzle depends on missing the crucial cognitive step. If you are auditorily aware only of sounds and their qualities, and if any consciousness of ordinary object and events is mediated by some further non-perceptual

cognitive states, then the puzzle dissolves apart from the question how it could strikingly seem that you are aware of commonplace objects and events in audition. Since the seeming requires explanation, even then a version of the puzzle persists.

If, however, the content of audition is very rich and audition represents, for instance, something like *train oncoming*, or *glass breaking*, in a way that involves exclusively perceptual states, then the puzzle is at its most pressing.

If the truth is somewhere in between, and audition furnishes awareness of more general categories like *source* or *object* or *event*, then the puzzle still arises. How could an extra-acoustic object or event be among the objects of auditory perceptual experience when sounds are in the first case the things we hear? How, in general, is it possible for an ordinary material thing or happening to be among the things we hear, when the immediate and proper objects of auditory perception are sounds? How, that is, could audition ever represent the presence of something that is not a sound?

3. The ‘composite snapshot’ conception of perceptual experience

I want to suggest that the puzzle just described ultimately has its source in a conception underwritten by visuocentrism in perceptual theorizing. Some explanation is needed.

It is fair to say that the traditional empiricist conception of overall perceptual experience is what we might call the ‘composite snapshot conception’ of experience, with an emphasis on ‘composite’.¹ The composite snapshot conception holds that perceptual experience is comprised of a set of discrete, modality-specific experiences superimposed (in the sense that each remains evident) to create one’s total perceptual experience at a time. According to this way of understanding perceptual experience, vision has a certain content characterized by colors and shapes, and, perhaps, ‘visual objects’; audition’s content comprises sounds, their pitches, and loudness; smell has a content characterized by odors and olfactory qualities; touch reveals textures, degrees of warmth, and pressure; and so on for each of the perceptual modalities. Compare David Lewis’s characterization of the ‘color mosaic’ conception of visual experience:

¹ Alva Noë's (2004) discussion of what he calls the 'snapshot conception' of experience is unconnected with, though it served as inspiration for, this way of thinking of perceptual experience.

Those in the traditions of British empiricism and introspectionist psychology hold that the content of visual experience is a sensuously given mosaic of color spots, together with a mass of interpretive judgments injected by the subject. (1966: 357)

Similar quality-mosaics could be described for each sensory modality according to the composite snapshot conception of perceptual experience. Each modality, according to this traditional empiricist picture, delivers a discrete snapshot of the world from its unique perspective that is distinct from each of the others. Vision could not share elements of audition's snapshot, and vice versa. The sum total of these snapshots, a sort of composite snapshot, constitutes and exhausts the content of one's total perceptual experience.

This traditional picture does not necessarily rule out that there could be 'common sensibles' accessible to more than one modality. For example, shape properties might be experienced through vision and through touch. What the traditional empiricist conception assumes is that these experiences are modality-specific and distinctive. There is, therefore, a distinctively visual way of experiencing shape that differs from the tactile experience of shape. This animates the long history of resistance to answering affirmatively Molyneux's question whether a subject without relevant background experience could visually identify a cube formerly only felt.

The traditional conception stems from thinking of the senses as distinct systems or channels of awareness. They are understood to involve separate processes, and to work in isolation from each other until some relatively late stage. Each modality, in addition, delivers an experience with a distinctive qualitative character that could not be created by any other modality, and furnishes only an experiential ingredient for one's total perceptual experience.

The traditional story is false in crucial respects and incomplete in others. I want to suggest that an important class of perceptual effects that has gone relatively unrecognized or unappreciated by philosophers gives us good reason to think that the composite snapshot conception of experience is incorrect. But the illusions that I shall discuss do not have merely negative implications. They provide the ingredients for the beginning of a solution to the puzzle about audition I described at the outset. They also illuminate

perception in perhaps its most significant respect and teach us what we could not otherwise have learned with attention restricted to vision, or to any other individual modality.

4. Cross-modal illusions

The perceptual effects I have in mind are ones in which what is sensed through one modality affects what is experienced in another. One familiar example, the ventriloquist illusion, has been well studied since the nineteenth century. Seeing the movements of a puppet's mouth affects where one hears the sound of a voice. But the effect is not limited to the perception of speech. Work in the second half of the twentieth century since Howard and Templeton (1966) has confirmed various ways in which the visual location of a stimulus affects perceived auditory location. The effect is neither cognitive nor inferential, but results from cross-modal perceptual interactions (see Bertelson 1999; Vroomen, Bertelson, and de Gelder 2001).

Cross-modal connections similarly are revealed in several other varieties of surprising illusion. Visual capture demonstrates that sight alters, for example, the tactile perception of object size and one's proprioceptive impression of bodily orientation: seeing a hand that is not one's own results in feeling one's hand to be where it is not (Hay, Pick, and Ikeda 1965; Pick, Warren, and Hay 1969), and seeing a larger object affects felt object size (Rock and Victor 1964).

In the fascinating McGurk effect, subjects shown video of a speaker articulating the velar (pronounced with the back of the tongue on the soft palate) /ga/ sound while presented with audio of the bilabial (pronounced with the lips) /ba/ sound experience an alveolar (pronounced with the tip of the tongue behind the teeth) /da/ sound (McGurk and MacDonald 1976). The McGurk effect takes place when conflicting auditory and visual information about speech are reconciled into a kind of 'average' or parsimonious percept.

Each of the preceding effects, however, could be explained in terms of vision's dominance over some other modality. Perhaps visuocentrism is vindicated by vision's dominance in perception over the other modalities?

Not so. Ladan Shams and her colleagues recently have discovered a class of illusions in which audition affects vision (Shams, Kamitani, and Shimojo 2000, 2002). In

the *sound-induced flash illusion* subjects presented with a single visual flash and double auditory beep have the same visual experience as when presented with a double visual flash accompanied by a double beep. They see two flashes instead of one when they hear two beeps -- the double auditory beep affects visual experience.

A single flash accompanied by multiple beeps is perceived as multiple flashes.

This phenomenon clearly demonstrates that sound can alter the visual percept qualitatively even when there is no ambiguity in the visual stimulus (2002: 152).

Three features of this result are significant. First, it is neither cognitive, nor inferential, nor based on some strategy adopted to respond to ambiguous or conflicting experiences.

Shams, Kamitani, and Shimojo (2002) maintain that audition influences the phenomenology of vision as a result of cross-modal perceptual interactions where we otherwise might expect vision and audition each to be autonomous and univocal.

Second, these and many other cross-modal effects are pre-attentional. ‘...Cross-modal interaction reorganizes the auditory-visual spatial scene on which selective attention later operates’ (Bertelson and deGelder 2004: 165). Finally, no semantic contribution from familiar bimodal contexts is necessary to generate the effect. It appears to be a perceptual effect grounded at a relatively low level. The effect does not result from learning for a particular context, and does not require specific bimodal experience. It is an audition-induced phenomenological change in the character of visual experience that persists through shifts in setting and stimulus characteristics. (Shams, Kamitani, and Shimojo 2002: 147)

5. Explaining cross-modal illusions

What are the consequences of cross-modal illusions for philosophical thinking about perception, perceptual content, and perceptual experience?

Cases in which one perceptual modality affects experience in another modality are familiar from the study of synaesthesia (see, for instance, Baron-Cohen 1997, Cytowic 1998 and 2002, and Harrison 2001). Some people systematically and persistently experience colors when hearing sounds, experience shapes as a result of taste, or experience sounds in terms of colors, shapes, or flavors. But these rare cases do not involve a special perceptual capacity, such as the ability to see into the ultra-violet range

or to detect the smell of cancer. Synaesthesia is something like a quirk in processing. Synaesthetes do not literally perceive the color of a sound or the shape of a taste, since sounds and tastes lack those colors and shapes. The experience always involves illusion.

The cross-modal illusions I have been considering are not the results of simple quirks of processing. Synaesthesia differs from ordinary cross-modal illusions in that information from one sense modality accidentally in synaesthesia, but not in ordinary cross-modal interactions, impacts experience in another. Cross-modal illusions, unlike synaesthetic ones, are intelligible responses to extra-ordinary situations. The responses are intelligible because they result from mechanisms that are perceptually effective and thus useful in overwhelmingly prevalent environmental conditions. The world does not frequently contain sounds at locations that differ from the visible locations of their sources. The size of a sphere one holds usually corresponds to the size of the sphere one sees oneself as holding. Acoustic events tend to match in number the salient visible events that are their potential sources.

Thus, such cross-modal effects demand a form of explanation that differs from that of synaesthetic illusions. To explain the influence of one modality upon what is experienced in another modality, in a way that captures the environmental significance of correlations across multiple modalities and, thus, their adaptive significance, requires appeal to some common factor that makes principles for grouping and organizing stimuli across the modalities intelligible. Such intelligibility and, thus, environmental and adaptive significance are absent from synaesthetic groupings.

These considerations are reflected in what have been called unity assumptions for (non-synaesthetic) cross-modal interactions (see Welch and Warren 1980). For example, when an incongruence (spatial or temporal) between stimuli from different modalities is relatively limited and when concordance surpasses some threshold, a common environmental source likely accounts for both stimuli. The perceptual system's response under these conditions exhibits cross-modal biases, recalibrations, or illusions.

Unity assumptions and the cross-modal illusions they explain illustrate that visual and auditory stimuli are treated as evidence of a single environmentally significant entity or event and that a perceptual 'unit' is formed according to principles analogous to those involved in Gestalt formation from vision and from audition. The principles according to

which perceptual units are formed in the cross-modal cases, however, are not limited to a single modality, but deal with the integration of information from different sensory systems. In order to avoid characterizing cross-modal illusions as accidental quirks of processing, these principles must be understood to invoke assumptions about a common environmental object or event that gives rise to sensory stimulation in multiple modalities, and about how it does so (see, for instance, de Gelder and Bertelson 2003, Wallace et al. 2004). Welch and Warren argued early in the study of cross-modal interactions that to explain such effects requires ‘the supposition that intersensory bias is a result of an attempt by the perceptual system to maintain a perceptual experience consonant with a unitary event’ (1980: 638). The important point is that these unity assumptions are not specific only to a particular modality; rather, they amount to either modality-independent or multi-modal assumptions concerning environment particulars that stimulate perceptual experience. They are, in effect, modality-independent assumptions about the sources of sensory stimulation. It is precisely because these grouping principles capture genuine regularities in the world of objects and events that awareness across different modalities constitutes genuine perceptual awareness of objects and events in the world.

But there is still a gap between influences across the modalities at the sub-perceptual level and the failure of the composite snapshot conception at the level of perceptual awareness. Sub-perceptual auditory processing might result in illusory visual experiences without this showing anything about either the content (its nature, constituents, or richness) of the overall perceptual experience or the appropriateness of the composite snapshot conception of experience. What is needed is a bridge between claims about the influence of one modality upon what is experienced in another and claims about the respective contents of each individual modality. I believe such a connection exists.

The grouping and binding principles I have mentioned appear systematically to affect or to determine modality-specific content. The question is whether assumptions concerning the unitary nature of an environmental stimulus to vision and audition merely causally determine perceptual content that remains distinct and discrete in each modality or whether they reveal that perceptual content includes unitary constituents shared among

perceptual modalities.

The perceptual system deploys principles designed to track, in a causally or counterfactually dependent way, the kinds of ordinary objects and events that lead to auditory and visual stimulation. But this assumes perception utilizes modality-independent or multi-modal characterizations of such objects and events. Describing these operations involves attributing to perception some traction on ordinary objects and events in a sense that goes beyond the modality-specific notions of ‘visual object’ or ‘auditory event’ deployed within a given modality. The idea is that experience is shaped by multi-modal organizing principles, and that such principles track objects and events, so perception itself involves a dimension of multi-modal or modality-independent content that cannot be characterized in purely auditory or purely visual terms. Such particulars, in addition, appear to serve as the locus for intermodal binding of perceptible qualities, so that the cube one holds is experienced as the very same particular one sees. The blip one hears is experienced to stem from the same event as the flash one sees. Perceptual experience therefore has a dimension of content that cannot be captured by a composite of modality-specific -- proper or unimodal -- snapshot-like contents.

Given that visual experience is affected and constrained by audition, and that auditory experience is affected and constrained by vision, it is therefore plausible to ascribe a dimension of content characterized in modality-independent or multi-modal terms even to vision and audition themselves. The very same amodal content might be shared by vision and audition. Visual flashes are visible as flashes of events in the environment; audible blips are audible as blips from environmental sources. It seems fair to suppose then that the object- or event-involving character of a given modality stems from underlying multi-modal principles and content with potential for sharing across modalities. But, even in the case of vision, such content cannot be captured by purely visual principles, and requires appeal to relations to audition and other modalities. Likewise, audition might involve a level of content that deals with environmental particulars and that it shares with vision.

We then have a foothold on the solution to the puzzle about audition I set out earlier. Audition has an object- or event-involving character because modality-independent or multi-modal principles shape auditory experience and ground a level of

content that cannot be characterized in purely auditory terms. We hear sources, objects, and events, and not just sounds, pitches, and timbres, because the senses do not act as isolated systems that deliver only neat modality-specific contents from which we learn to infer the presence of ordinary objects and events.

I am not suggesting that the modalities of sense perception cannot be differentiated. Distinct perceptual modalities might each furnish exclusive awareness of a range of proper sensibles. Proper sensibles could include qualities such as color, timbre, taste and warmth, and particulars such as odors and sounds. I do not wish to eliminate all boundaries among the senses.

What I am suggesting is that a convincing explanation of the cross-modal effects requires appeal to a dimension of perceptual content shared across the modalities. If this is right, then any snapshot that emerges within a specific modality is itself already a multi-modal sculpture infused with information shaped by and gleaned from the other modalities. Even the content of vision itself cannot thoroughly be understood in complete isolation from the other modalities.

Not only does the traditional empiricist conception that likens perceptual experience to a composite of discrete modality-specific snapshots fail as a characterization of perceptual experience, but its failure reveals a perilous flaw in the visuocentric thinking from which it stems. The tendency to take vision as an independent and representative paradigm for theorizing about perception is not just incomplete, but the approach to theorizing about perception it fosters encourages us to consider each modality as an autonomous mode of awareness and domain of inquiry. I have aimed to show that this undermines a complete understanding of perception and leaves out what is most critical for resolving long-standing philosophical debates about the possibility and grounds of perceptual access to a world of things and events. Comprehending the relationships among modalities is helpful in resolving puzzles about audition's object- and event-involving character. It is, however, essential to any satisfactory philosophical understanding of perception. The tyranny of the visual threatens to blind us to the nature and character of perceptual awareness.