Object Perception: Vision and Audition

Casey O'Callaghan*
Rice University

Abstract

Vision has been the primary focus of naturalistic philosophical research concerning perception and perceptual experience. Guided by visual experience and vision science, many philosophers have focused upon theoretical issues dealing with the perception of objects. Recently, however, hearing researchers have discussed *auditory objects*. I present the case for object perception in vision, and argue that an analog of object perception occurs in auditory perception. I propose a notion of an auditory object that is stronger than just that of an intentional object of audition, but that does not identify auditory objects with the ordinary material objects we see.

1. Objects in Perception

Humans understand the world in terms of objects. We take the environment to be populated by things like forks and bottles and steaks. Whether or not the world contains any such items, medium-sized dry goods are one central component to our conceptual schemes.

Objects also feature in how we perceive the world to be. Several initial considerations support this intuitive claim. Birds attract and hold our attention as we track them in flight. So, if we can attend only to what we perceive, we visually perceive material objects. Objects are subjects of empirical beliefs formed on the strength of visual experience. If perceptual experiences constrain the structure and content of thoughts, the experience of objects explains the structure and content of thoughts about objects. Common actions, such as reaching for a spatula or swinging a racket at a ball, target objects. If the details of what we do in such cases depend upon characteristics of what we see, perception targets objects. Thus, attention, perceptual belief, and action hint that experience not merely causes cognition aimed at objects, but that it does so because objects figure among the things we perceive. An object-involving structure for perception helps explain the object-involving structures of attention, belief, and action.

This much seems compelling, at least when we focus upon vision. Seeing, however, typically is presented as an exemplar of perceiving. Given the prominence of objects in visual perception, it is tempting to think that all perceiving concerns objects, their features, and their

arrangement. Audition, touch, olfaction, and gustation thus may follow the model of vision's organization, character, and function. According to this line of thought, the various sense modalities involve phenomenologically distinctive ways of becoming acquainted with objects. Armstrong says, simply, 'In perception, properties and relations are attributed to objects' (20). Frequently, this conception is apparent in discussions of the various sensible qualities. Shoemaker, for instance, says we experience sensible qualities 'as belonging to objects in our external environment – the *apple* is experienced as red, the *rose* as fragrant, the *lemon* as sour' (97; qtd. in Matthen, *Seeing* 288). Pasnau maintains that the sounds we hear are properties attributed to objects such as bells, whistles, and sirens. Perceiving ordinary material objects and their features, in the commonplace view, extends as a rule to modalities apart from vision.

Reading lessons about other perceptual modalities off of vision is poor methodology in science and in philosophy. We should not assume that the structure and function of auditory, gustatory, or olfactory awareness mirrors that of vision. Doing so risks neglecting the diversity that is most striking about experience across the modalities. Any hope for a comprehensive naturalistic theory of perceiving depends upon a close examination of non-visual modalities.

Do all modes of perceiving aim at objects? Consider touch. With noteworthy exceptions, such as rainbows, shadows, and beams of light, touch does seem to reveal many of the same objects we see. Olfaction, however, differs. Though we smell smells or odors, olfaction does not obviously involve awareness *as of* rose bushes or patchouli plants, which may be gone before their smells. Similarly, audition involves awareness of sounds, but fails to guarantee awareness *as of* the material objects that make them. Sounds are unlike ordinary tables and chairs – you cannot grasp or trace a sound – and sounds are not heard to be properties or qualities of tables and chairs, since sounds do not seem bound to ordinary objects in the way that their colors, shapes, and textures do (O'Callaghan ch. 2). Auditory experience presents sounds as independent from ordinary material things, in a way that visual and tactual features are not. This presents a nice *prima facie* case that not all perception involves object perception.

Surprisingly, then, researchers recently have extended discussions of object perception beyond vision and touch to other modalities. In particular, the auditory perception of objects has come into focus. The objects in question, however, are not ordinary material objects, but *auditory objects*. This raises three pressing questions for perceptual theorizing. First, given the disanalogy already detailed, what grounds the claim that audition, like vision, involves a form of object perception? Second, what are auditory objects? If they are not ordinary material objects, what about their individuation and identity conditions warrants calling them 'objects' at all? Finally, does the perception of objects in audition vindicate some form

of the visuocentric claim that all perceiving involves the perception of objects and their features?

In this article, I hope to answer these questions. I begin by presenting the theoretical and empirical case for the claim that humans perceive objects. I focus in sections 2 and 3 upon vision and the objects we see. The evidence supports the claim that we see not just qualities or features, but individuals that bear them. Furthermore, the individuals we see include not just locations or surfaces, but spatio-temporally continuous objects that correspond to familiar, material objects. Developing an explanatory account of object perception and determining whether and how it extends to audition requires that we distinguish among a number of different conceptions of an object, including those of *ordinary object*, *material object*, *intentional object*, *proper object*, *visual object*, *auditory object*, and, more generally, *perceptual object*. With these distinctions in view, I turn in section 4 to audition and present the case that hearing involves a form of object perception, in a sense stronger than that audition has intentional objects or proper objects. Audition involves awareness not just of qualities, but of individuals. Furthermore, it involves awareness as of a variety of individuals that deserve the name 'auditory object' in light of their composition and continuity. Auditory objects, like visual objects, are mereologically complex individuals that persist through time. Such objects, however, differ in critical respects from the ordinary material objects we see. Most notably, the mereology according to which they are perceptually individuated and identified is primarily temporal rather than spatial. I discuss in section 5 the sense in which vision and audition nonetheless both count as forms of object perception, though this sense is quite different from what was canvassed at the outset. This discussion delivers a broader understanding of a perceptual object and reveals an important lesson concerning the structure of perceptual experience. Each should serve as impetus to future research in the naturalistic philosophy of perception.

2. The Case for Objects

Vision provides the best case for the claim that humans perceive objects, and vision has been the focus of debates about object perception in philosophy and cognitive science. Perhaps surprisingly, skepticism about object perception until recently has been prevalent. An infamous quote demonstrates the sentiment.

> Perceptual systems do not package the world into units. . . . The parsing of the world into things may point to the essence of thought and to its essential distinction from perception. (Spelke, 'Where Perceiving Ends' 229; qtd. in Scholl 2)

Against the claim that perception parses a scene into objects, one might contend that perceptual systems represent or reveal, at most, a scene's

Fig. 1. Parsing ordinary objects.

qualities, features, locations, or surfaces. Skeptics offer reasons to stop some distance short of attributing to perception a grasp upon objects.

The barriers concern how we individuate objects at a time and how we identify objects over time. We count or parse objects in idiosyncratic ways. The pushbuttons belong to my telephone because they are attached, but the pens clinging to the aluminum can on my desk, though attached, do not belong to it. The fob and split ring comprise the keychain, but the keys do not. Why don't we take the keys and ring to comprise an object that excludes the fob (see Fig. 1)? Performance that exhibits our mastery of the answer invokes conceptual capacities regarding keychains, fobs, and keys. If automatic sensory processes with no access to such information drive perceptual systems in a bottom-up manner, perception cannot parse a scene into such objects. In addition, what makes something the kind of object it is often depends not upon visible features such as color and shape, but upon properties that are hidden from view or imperceptible. Nothing visible differentiates a fruit, a wax fruit, and an autostereoscopic image of a fruit. Vision, furthermore, seems unequipped to grasp the complex survival conditions and modal properties required to individuate ordinary objects that persist and change. A thing's parts might be replaced, or it might be completely disassembled and reassembled. One object might split into two, and bits of different things might fuse. A hunk of clay might survive being smushed, while the statue it constitutes does not.

If, however, perceptual experience presents only a two- or three-dimensional mosaic of qualities or sensible features that evolves over time, or if it reveals an unbroken arrangement of surfaces akin to Marr's 2½-D

sketch, then extra-perceptual cognition might inject further interpretative judgments required to grasp and count objects.

Notice, however, that two different kinds of worry are in play. The first concern is whether perception could individuate and identify objects in ways that correspond to sortal concepts deployed in common thought about objects. The second concern is whether perception could carve the world into object-like units or whether it might serve the needs of attention, thought, and action with more modest resources, such as features, locations, or surfaces. So we must distinguish the question whether perception captures the nuanced conceptual structure distinctive to thought and talk about common sorts of objects from whether there is some more basic or generic notion of *object* applicable to visual perception. We might answer 'yes' to the second even if we answer 'no' to the first.

Why believe vision carves the world into units? Sensory systems detect many features and qualities. Vision detects colors, textures, patterns, orientation, and motion, among other features, and does so with dedicated resources – single cells literally respond selectively to the presence of certain features (see Barlow). Suppose, then, that vision just detects features like *redness*, *squareness*, and *roughness*. A *feature-detection* model of vision is simple and physiologically tractable. But it fails. Consider seeing a red square beside a blue circle. Now consider seeing a blue square beside a red circle. The two experiences differ. If perceiving is detecting features alone, however, the two are equivalent experiences – of redness, blueness, squareness, and circularity. Nothing on the feature detection account non-circularly explains the *coinstantiation* of a color and a shape. Suppose we add locations, such as *leftness* and *rightness*, to the list of features detected. This is no help, since we are left with equivalent experiences of redness, blueness, squareness, circularity, leftness, and rightness. Again, nothing captures, without regress, that a color, shape, and location are coinstantiated.

This problem, dubbed the 'Many Properties' problem (see Jackson; Clark for extensive discussion) resembles the problem of determining which features belong together and must be *bound*. Explaining feature binding requires explaining coinstantiation. Explaining coinstantiation requires a predicative mechanism capable of attributing features in groups to common items. This requires distinguishing features from that to which they are attributed. In short, it calls for *sensory individuals* (Cohen).

Locations, construed as *bearers* of visual features, solve the Many Properties problem. If distinct locations are distinct sensory individuals that instantiate visual features, then vision may attribute redness and squareness to one place while locating blueness and circularly at another.

The *feature-placing* model solves the Many Properties problem because it admits units or individuals, but it does not go far enough. We are able visually to track individuals that move – I experience a single individual as I watch a blue spot travel. Locations, however, do not change location.

Furthermore, I experience perceptible individuals to survive changes to certain properties and not others. I might perceive a given individual to survive a gradual color change, but not a dramatic one. A location, however, survives both. An individual may perceptually seem to cease entirely to exist, while a location occupied later by a new feature is the same individual. Finally, adjoining locations often are experienced as parts of the same individual – they comprise parts of a *region* the whole of which instantiates some visible feature, such as shape or uniform color. That region might perceptibly survive deformation, or change to the locations that comprise it.

Admitting *surfaces* among perceptible individuals avoids many shortfalls of locations alone. Surfaces are depthless regions that bear sensible features. Surfaces can deform and translate through space. Surfaces survive certain changes to qualities and features but not others.

What kinds of surfaces do perceptual experiences track? Spelke ('Where Perceiving Ends' 229) suggests, 'Perceptual systems bring knowledge of an unbroken surface layout'. Taken strictly, such a continuous visual lamina cannot explain the phenomenology attending bistable figures such as Fig. 2. Such Gestalt shifts result from different ways to distinguish a figure from its background, which requires detecting *edges* that mark distinct *bounded* visible regions. Since a figure perceptually seems to rest in front of its background, its visible edges must mark its own boundaries.

The perceptual parsing of a scene, therefore, requires recognizing units that correspond to visible regions. Such units, however, may extend beyond the parts we see to possess visible qualities. Perceptual systems parse surfaces in a way that recognizes that surfaces may continue uninterrupted behind an occluding surface. Distinguishing figure from ground, for instance, involves not just detecting visible borders between regions, but also allocating edges exclusively to a single surface region. The figure region 'owns' a common edge, which thus forms its boundary, while the background continues behind that boundary. Thus, in Fig. 3a, regions x and y are parsed as belonging to a single surface that completes behind an occluding surface, while, in Fig. 3b, x and y appear to be separate surfaces. Enhanced performance on search and recognition tasks demonstrates the impact of such *amodal completion* upon how a scene is perceptually parsed. Principles of amodal perceptual completion help you to see a runner behind a barrier in Fig. 4a, but make this difficult when identical runner parts are presented in strips, as in Fig. 4b. Figure 5 also illustrates how amodal completion aids recognition. Sometimes, completion leads to surprising illusions. Figure 6 appears to depict a triangle occluding three circles. The illusory triangle even appears to contrast its background in brightness. Perceptually parsing a scene thus involves detecting and representing information about surfaces that are arranged in three-dimensional space and that may have hidden parts.

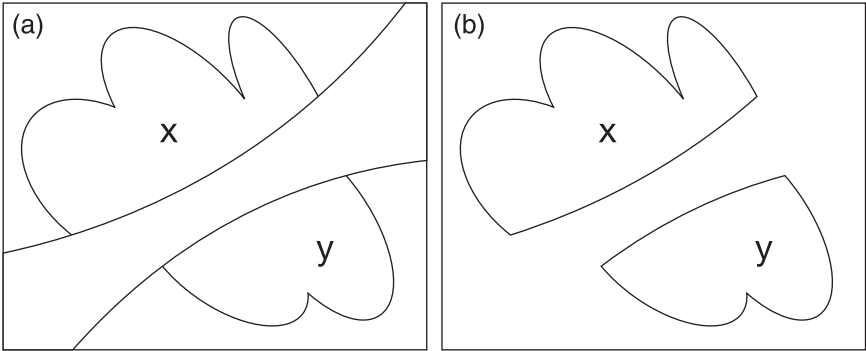Fig. 2. Bistable figures and figure-ground distinctions.

Fig. 3. Occlusion. (a) Regions *x* and *y* appear to belong to a single partly occluded surface. (b) Regions *x* and *y* belong to visibly distinct surfaces.
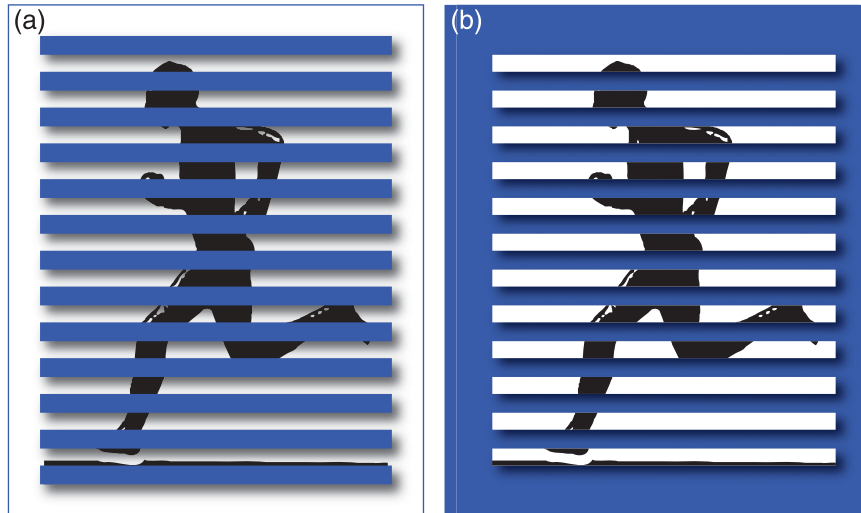
Fig. 4. Occlusion and recognition. (a) The runner is easily identified as single connected object behind the blue occluding strips. (b) The runner is difficult to parse as single connected object when same parts are presented as coded in front of a blue surface.
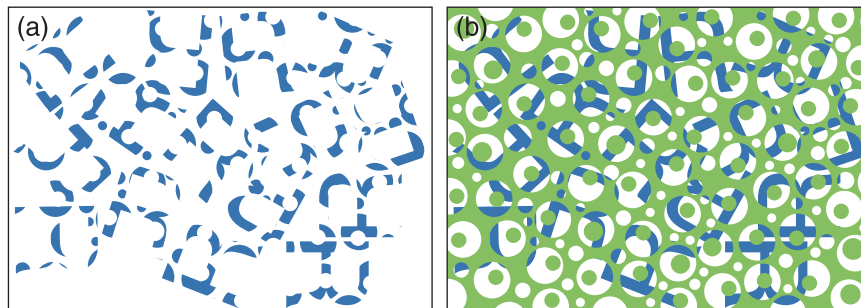

Fig. 5. Occlusion and recognition. (a) Apparently random blue surface segments. (b) The blue segments are more easily identified as parts of the letter 'R' when they appear to be behind the green occluder.

How could perceptual systems track partly occluded surfaces? Vision relies upon 'T-junctions' where edges meet (as present in Fig. 3a, but absent in 3b) and stereoscopic depth cues (see Fig. 7) in order to determine the mereology and spatial arrangement of amodally completing surfaces. Vision, in effect, invokes a subpersonal grasp upon principles of occlusion, such as that regions that do not own borders may complete behind a border owner to form a single surface (see Nakayama et al. for review).

Surfaces so understood capture much of what we want from a theory of perceptible individuals or units. They nonetheless fail to explain further perceptual phenomena, which in turn suggest that surface perception
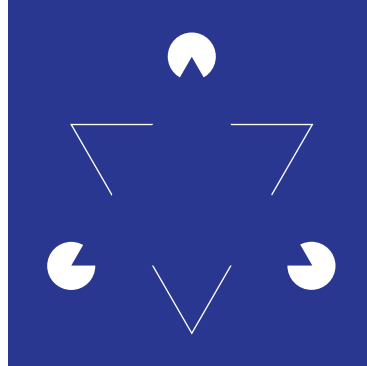
Fig. 6. Kanizsa triangle.

subserves object perception. First, perceptual systems individuate and track groups of connected surfaces as belonging to a single individual or unit. They also track visibly disconnected components that, due to their shared pattern of motion, appear to comprise a single individual, as in Fig. 8 and Supplementary Video S1. Furthermore, we perceptually represent and track individuals through time despite seeing entirely different surfaces, as when an object rotates. Similarly, perceptual constancies concerning three-dimensional shape, for instance, manifest a perceptual grasp upon individuals that comprise multiple surfaces and persist despite perspectival changes to their appearances. These phenomena indicate that bounded, connected, cohesively moving three-dimensional constructions from surfaces feature in the organization of visual experience. Explaining vision requires recognizing objects.

3. What Is a Visual Object?

Empirical theories of vision strive to correctly capture how perceptual systems individuate objects at a time and how they identify and track objects over time, given information delivered by early sensory processes. Vision also must serve the needs of object recognition with the help of long-term memory, which guides categorization and concept application. The intermediate visual level thus need not represent objects as belonging to full-fledged everyday concepts, but might offer information concerning the basic spatio-temporal characteristics of something akin to material objects, which then drives recognition of such objects as belonging to familiar kinds. Thus, visual object individuation requires capturing the spatio-temporal characteristics of objects (cf. Spelke, 'Principles of Object Perception').

Vision individuates objects *at a time* primarily according to spatial criteria, such as continuity, contact, and boundedness. Features such as color, texture, and shape play a less critical role, except when spatial information
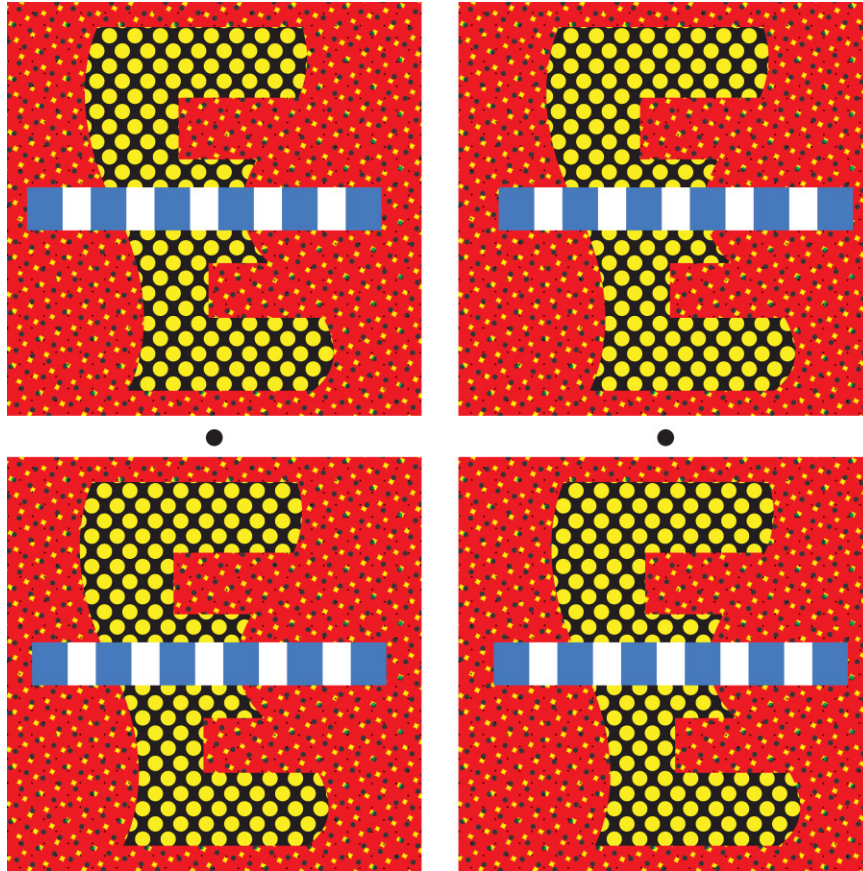
Fig. 7. Depth cues, occlusion, and parsing. Cross your eyes until the two black dots visually coincide and stereoscopically fuse the images. The top figure appears as a single 'E' partly masked by the occluding bar, which is coded in front due to binocular cues. The bottom figure appears as two distinct halves of an 'E' that hover in front of the bar, which is coded in back due to binocular cues.
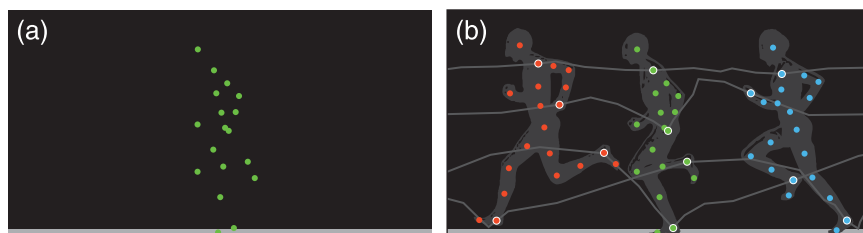

Fig. 8. Motion cues to objecthood. (a) A group of green dots. (b) Putting the dots in motion provides cues to a continuous moving object. (c) Supplementary Video Clip S1. <http://www.blackwell-compass.com/home_video#phco1>.

is ambiguous (see, e.g., Leslie et al.). Whether bounded surfaces share edges or are in contact is more critical to object individuation than uniform color, texture, or shape.

Time, however, resolves many of a static display's spatial and structural ambiguities. Changes both to one's perspective and to the scene deliver information that helps to individuate objects, for instance when two surface regions appear to touch only from a certain viewing angle, as in Fig. 9. Visible cohesion through motion and even coherent patterns of motion among disconnected elements are guides to individual objects (as in Fig. 8).

Notice, however, that using motion to individuate objects requires a more fundamental capacity to identify or track something as the same from moment to moment. The well-known phi phenomenon nicely illustrates this capacity (Wertheimer). When shown a dot, followed by a gap of around 2 seconds, and then a second dot some distance from the first, subjects experience two separate dots. When the gap is narrow, say one-half second, subjects experience a single dot to move from left to right. The effect persists when the dots differ in color or when the presentation uses different shapes – subjects then report the dot to change
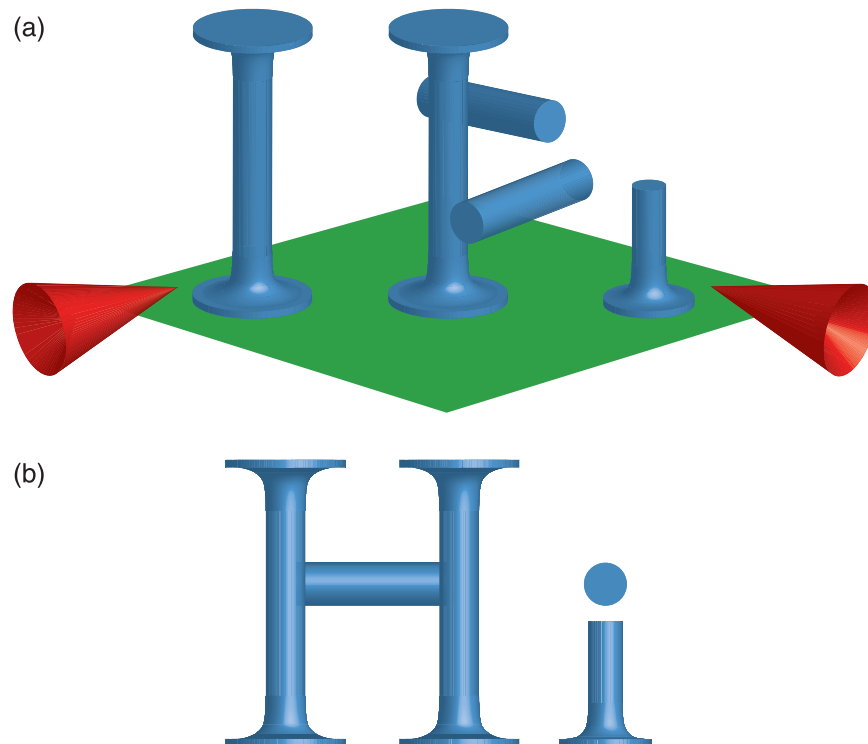
(a)

(b)

Fig. 9. Resolving structure through perspective. (a) Viewing this display from the perspective of either of the red cones reveals the structure depicted in (b).

color or the item to change shape as it moves. The phi phenomenon persists even when it involves a bunny and a duck – the bunny appears to change into a duck as it moves. Vision represents a single individual to persist, move, and change features when the delay is short; it represents distinct individuals in distinct locations when the gap is long.

We can, in addition, simultaneously keep track of distinct individuals over time. Subjects who view a display in which two items move diagonally to opposite corners of a screen most frequently see two individuals to *stream*, or travel past each other on straight paths, though the display is compatible with items that collide and rebound, or *bounce* (see Fig. 10 and Supplementary Video Clip S2). Subjects prefer straight-line motion, even when the streaming percept does not minimize visible change. Some abrupt feature changes, such as those that accompany swapping color and shape, are even preferred over the bouncing percept. Features do sometimes impact how objects are tracked over time, but do so primarily when they are salient or when change is dramatic, and when spatio-temporal information is ambiguous (see Feldman and Tremoulet). These kinds of results support the claim that vision tracks objects through time, in the first instance, according to spatio-temporal continuity.

A dramatic demonstration of this, and of the early visual processes responsible, involves multiple object tracking (MOT) tasks (see Pylyshyn, 'Visual Indexes'; *Things and Places* for discussion). In MOT experiments (see Fig. 11 and Supplementary Video Clip S3), subjects view a display that contains about eight dots. Four to five dots then flash briefly to mark them as targets. The dots then all begin to move around (perhaps even appearing to disappear momentarily behind occluders). Subjects, remarkably, are able correctly to identify the original targets after many seconds of motion; vision successfully tracks the targets. This capacity tellingly trails off when more than around four targets are present. Primarily on the strength of MOT demonstrations, researchers argue that early vision assigns a fixed
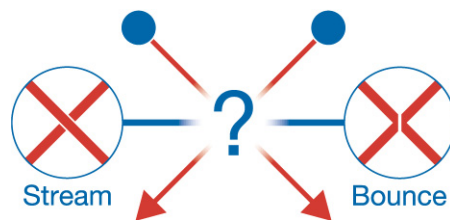

Fig. 10. Streaming and bouncing. (a) The motion paths of two visible dots that move toward each other, briefly coincide, then move apart are ambiguous between straight-line motion and collision with rebound – streaming versus bouncing. Adding an auditory stimulus at the time of coincidence increases the incidence of visual bouncing percepts (Sekuler et al.). (b) Supplementary Video Clip S2. <http://blackwell-compass.com/home_video#phco2>. Play the animation with sound muted, then with sound, to demonstrate the visual ambiguity between streaming and bouncing, and the impact of auditory cues. Animation appears with permission, courtesy of Robert Sekuler <http://people.brandeis.edu/~sekuler/AVinteractionDemo/victoriaAVdemo.mov>.
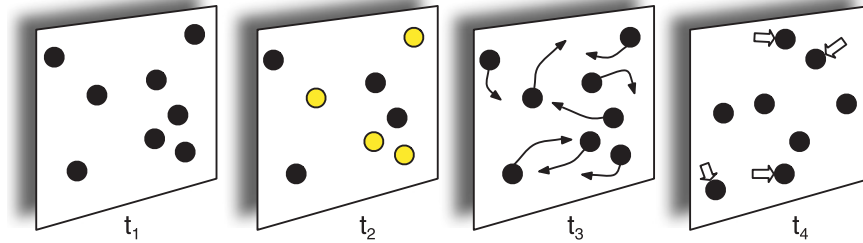
Fig. 11. Multiple object tracking. Observers successfully track up to four or five visible targets over time. Image and animation appear with permission, courtesy of Zenon Pylyshyn and Brian Scholl <http://ruccs.rutgers.edu/finstlab/MOT-movies/MOT-Occ-baseline.mov>. See also Supplementary Video Clip S3. <http://www.blackwell-compass.com/home_video#phco3>.

number of primitive markers or pointers that refer deictically to visual objects without representing their features. Such pre-conceptual 'sticky' object indexes ('FINSTs', for Pylyshyn) are assigned to individuals based on spatio–temporal criteria and are responsible for the capacity visually to track multiple items over time. Feature information later may be bound to such indexes to form full-fledged visual object representations (see also Leslie et al.).

The lesson of these examples is that perceptual mechanisms that shape visual experience carve out and track object-like individuals. As a consequence, such individuals figure prominently in the structure of many visual experiences by serving as the locus for the binding of visible features. Visual perception, then, is not merely a matter of projecting qualities from sensory receptors or single cells to higher cognition, and visual experience does not involve just an unbroken color array. Rather, vision involves extracting from sensory stimulation information about objects and their features. In what sense are such individuals object-like? They are spatially bounded, connected, or unified; they travel upon spatially continuous paths; and they persist in temporally continuous trajectories. Furthermore, they appear to be present at each given moment at which they exist, while recognition is governed by features at a time. Though they are seen to have spatial structure, they do not visually appear to have temporal parts.

Are the objects vision discerns and tracks *ordinary objects*? The bounded, cohesive individuals tracked by vision correspond for the most part to common items like tables and chairs. However, vision permits objects we would not recognize as ordinary, such as a ping–pong ball glued to a fork. And vision misses ordinary objects like the keychain, which comprises fob and ring but not keys. Vision also tracks individuals through kind differences, such as that reported by the call, 'It's a bird, it's a plane, it's Superman!' that ordinary objects may not survive (the life of a butterfly is an illustrative exception). Vision indeed may represent something as an object without representing any familiar kinds to which it belongs – vision need not represent something as a blender or an appliance to
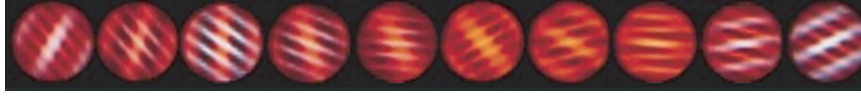
Fig. 12. Superimposed Gabor patches. Reprinted by permission from Macmillan Publishers Ltd: *Nature* (Blaser et al.), copyright 2000.

represent it as an object.[1] Vision's objects therefore need not be grasped as ordinary objects.

Perhaps vision represents *material objects* as such. Vision, however, treats rainbows, holes, and beams of light as objects, though none is a material object. Does vision misrepresent such things as material objects, or does it fail to grasp material objects as such? The answer turns on what is required to represent an object as material. Visual objects in fact may disobey plausible requisite principles. For instance, Blaser et al. claim that visual attention can select and track distinct but spatiotemporally coincident objects. Two superimposed Gabor patches, such as those depicted in Fig. 12, may provoke distinct visual object indexes assigned to individuals that occupy the same region. Furthermore, Leslie reports a surprising illusion in which we visually experience two solid items as passing through each other (see Fig. 13, the Pulfrich double pendulum illusion; Wilson and Robinson). Vision, Leslie concludes, does not always represent its objects to obey the principles of solid material items. 'This suggests that the visual system is really rather happy with the idea of solid objects passing through one another' (199). In light of this, vision's perceptual objects perhaps are best understood as Scholl and Pylyshyn suggest: 'There is a notion of a *visual object* that has been widely used to refer to visually primitive punctate spatiotemporal clusters' (26). Alternatively, Matthen's characterization of a visual object as a 'spatio-temporally confined and continuous entity that can move and take its features with it' (*Seeing* 281) does not rule out light beams, coincidence, or permeability.

Visual objects nonetheless correspond for the most part to material objects, and vision most likely evolved to represent material objects. Object representations drawn from vision may require supplementation by perceptual representations derived from other modalities, such as touch, to furnish our full perceptual understanding of material objects. Still further resources, such as a theory of space and matter, may inform our conception of material objects. But it is reasonable to hold that vision functions to reveal material objects in some of their guises, even granting exceptions.

4. Auditory Objects

I have claimed that we see objects. In many cases, visual objects correspond to material objects or medium-sized dry goods. It makes little
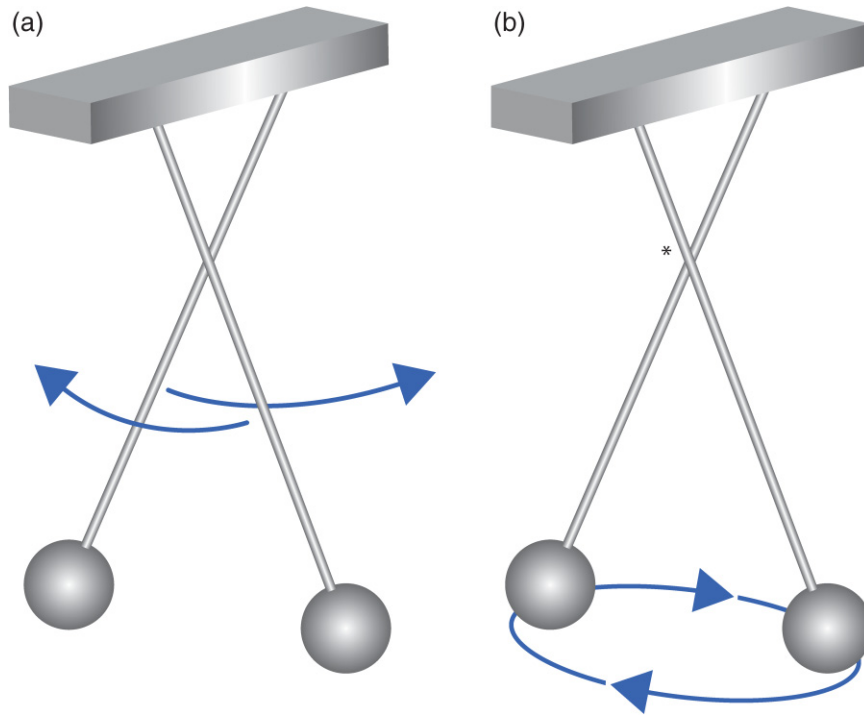
Fig. 13. Pulfrich double pendulum illusion. (a) The two arms of the pendulum in reality swing past each other on separate planes. (b) The two arms, however, visually seem to rotate around each other, which would require the solid arms to pass through one another.

sense, on the other hand, to suggest that we *hear* bounded, connected, cohesively moving three-dimensional constructions from surfaces *as such*. Audition is spatial, but I have never heard the boundary between two surfaces, nor have I heard something to complete spatially (at a time) behind an occluder.[2] Instead, it seems that I hear sounds, and that sounds give clues to ordinary material objects. But the individuation and identity conditions for sounds differ from those of material objects – sounds need not even correspond to objects. Audition's organization and structure does not, in the first instance, feature ordinary or material objects in the manner of vision. What, then, could researchers mean when they help themselves to the notion of an *auditory object* (see, for instance, Kubovy and Van Valkenburg; Scholl; Griffiths and Warren; Matthen, 'Diversity of Auditory Objects')?

It is worth making two things clear from the start. First, talk of auditory objects is not just a confusion or shorthand stemming from thinking of audition's *intentional objects*, which may not include objects in the familiar sense at all. If an intentional object of a perceptual state is something that

state concerns or represents, or at which it is directed, then the intentional objects of a perceptual state might include things apart from ordinary or material objects. If you perceive the black dog sitting on the bar, among your perceptual state's intentional objects may be a dog, the bar, a color, the relation of sitting upon, and perhaps the state of there being a black dog sitting on the bar. Audition's intentional objects thus might include sounds; instances of audible properties and relations such as pitches, octaves, timbres, loudness, and durations; and perhaps the state of affairs of one sound's being an octave higher than another. The notion of 'object' in 'intentional object' is more like that in 'direct object' than in 'material object'. Just as the claim that we see objects is not merely the claim that vision has intentional objects, talk about auditory objects is not merely talk about audition's intentional objects.

Second, talk about auditory objects is not just talk about the *proper objects* of audition. A perceptual modality's proper objects are intentional objects inaccessible by other perceptual means. Sounds are proper objects of audition, odors of olfaction, and colors of vision. Ordinary objects, however, are not proper objects of vision since they can be touched. Talk of auditory objects is not just talk of audition's proper objects, since it is an open question whether auditory objects are accessible to other modalities. The analogy between visual objects and auditory objects is intended to be much stronger.

The analogy, in fact, is based upon an intriguing similarity between audition and vision. The similarity is not that both vision and audition concern and ascribe features to ordinary or material objects, just as it is not merely that both vision and audition have intentional objects. Rather, the structure of perceptual experience in vision and in audition suggests that a more general notion of *perceptual object* captures a critical aspect of how perceptual experience is organized and draws attention to one central task common to both vision and audition. This parallel suggests a more viable sense in which vision and audition both count as forms of object perception, according to which perceiving objects involves individuating and tracking mereologically complex individuals.

Consider the case for auditory objects. When hearing, we perceive audible *features* such as pitch, timbre, and loudness. At a cocktail party, however, you are able to discern the sound of a familiar voice from across the room amid an array of other voices. Right now, I can hear Emily's voice downstairs while I hear the sounds of cars passing outside and a banging sound across the street. We are able, at a given time, to hear distinct things. You might hear the loud, high-pitched sound of the nearby trumpet while hearing the soft, low-pitched drone of the cement truck in the distance. Such an experience, however, cannot be captured in terms of mere feature awareness. An analog of the Many Properties problem exists for audition. As with vision, characterizing the sense in which groups of audible features qualify common items requires appeal to perceptible *individuals*, which, in turn, are necessary to explain our capacity simultaneously to

discern distinct sounds. A nearby audible individual might be high-pitched and loud while another distant audible individual is low-pitched and soft.

What, from the point of view of audition, is such an individual? As with vision, locations do not suffice, since an audible individual can travel from one location to another. Though auditory individuals audibly appear to have spatial locations, however, they are not represented to have spatial parts or complexity, and exhibit no spatial opacity. Surfaces, therefore, are not apt. In general, space is not nearly so critical to individuating auditory individuals as it is to discerning visual individuals. Vision individuates surfaces and objects primarily in virtue of spatial boundaries, but spatial features may be neither sufficient nor necessary for the individuation of distinct auditory individuals. When two separate speakers play different notes, audition ordinarily parses the auditory scene as two distinct audible individuals. However, when two separate speakers play the same note, we commonly hear a single audible individual. When a single speaker plays different notes simultaneously, we often hear two distinct audible individuals (see Fig. 14). Space, nonetheless, may aid in attending to distinct auditory individuals. Though it is very difficult to discern two different bird songs played from a single speaker, separating the signals and playing them through different speakers dramatically enhances one's ability to hear the two songs as distinct (Best et al.).

Consider, however, the role of time and temporal characteristics in audition. Audition researcher Albert Bregman has called the problem of
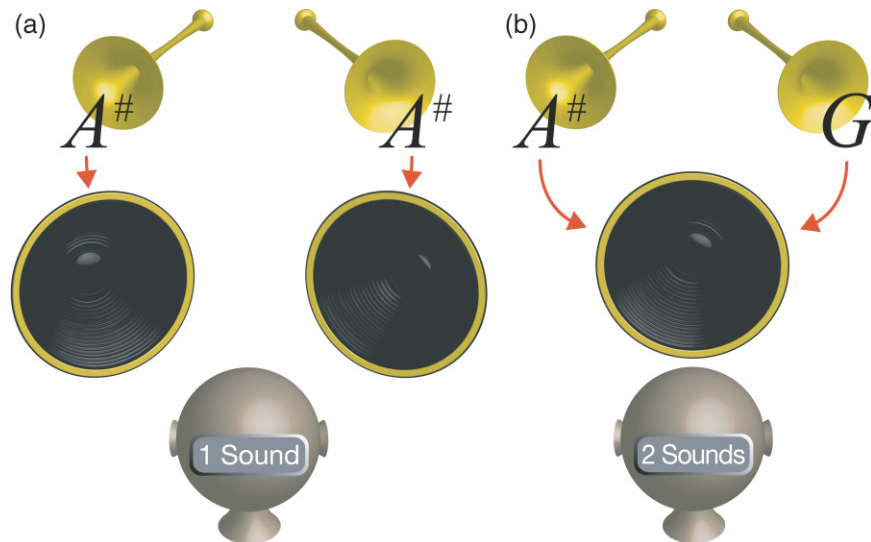

Fig. 14. Auditory objects and space. (a) Sounds at the same pitch from different locations may auditorily appear to comprise a single audible individual. (b) Sounds at different pitches from the same location may auditorily appear as distinct audible individuals.

discerning information concerning audible individuals and their features from complex wave information that arrives at the ears the problem of *auditory scene analysis*. Bregman likens auditory scene analysis to determining the number, location, size, and activity of boats on a lake by observing just the waves traveling up two small channels dug at the lake's edge. Audition, according to Bregman, carves the auditory scene into distinct *auditory streams* on the basis of stimulation by pressure waves. Auditory streams have qualities like pitch, timbre, and loudness *at a time*, and multiple auditory streams might occur simultaneously. You might hear a stream at high pitch while hearing a distinct stream at low pitch. Auditory streams thus are perceptual individuals that bear audible qualities. Auditory streams, however, also *persist through time* and survive changes to their audible qualities. A single stream might begin high–pitched and loud and gradually become low–pitched and soft while remaining the same audible individual. Indeed, since the identities of many recognizable sounds, such as the sounds of spoken words, police sirens, bird calls, and tunes, are tied to distinctive patterns of change in audible qualities through time, the constraints imposed upon auditory perception by the needs of recognition and concept application require that audible individuals exhibit temporal structure.

Auditory processes, in fact, determine the organization of streams in time according to principles that parallel how vision determines the constitution and arrangement of objects in space. An auditory analog of *edges* and *boundaries* exists in time, for example. While vision exclusively allocates *spatial* boundaries, audition exclusively allocates a *temporal* boundary to a single auditory stream. Due to the principle of exclusive allocation, a sequence of tones, *p p p q r r r* (see Fig. 15), may auditorily appear either as *p-p-p-q* and *r-r-r*, or as *p-p-p* and *q-r-r-r*, depending on the relative pitch distance between *p*, *q*, and *r* (cf. Bregman 14–15). When two streams are distinguished, the *q* must belong to one stream or the other.

Furthermore, the figure–ground distinctions and shifts expected to accompany such boundaries have an auditory analog in time. For instance, the sequence of tones, *p p p q r s q p p p* is heard as *p-p-p-p-p-p* and *q-r-s-q* when the pitch distance between *p* and *q* is great (see Fig. 16). Auditory attention may select one such item or stream as figure while the other becomes ground. This, notably, leads to difficulty discerning the temporal order of elements within streams. Attending to the stream, *p-p-p-p-p-p*, for instance, makes it difficult to report the order of *r* and *s* in *q-r-s-q* (Bregman 15). Furthermore, the temporal order of two discrete streams (such as those depicted in Fig. 18) often is mistakenly reported (either as *p-q–r–s t-u–v–w* or as *t-u–v-w p-q–r–s*) due to figure–ground effects (18). Such effects require selective attention that operates over distinct auditory individuals.

Not only does audition individuate distinct streams that persist through time, it identifies such streams as persisting despite the presence of masking noise. When a horn honks during a conversation, the conversation's earlier
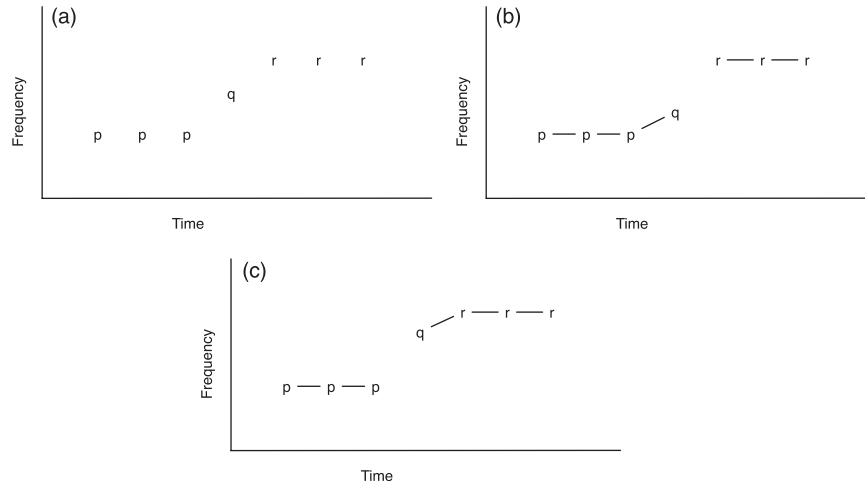
Fig. 15. Auditory streams and exclusive allocation. A sequence of tones (a) may be parsed either as in (b) or as in (c). Each tone is exclusively allocated to a single stream.
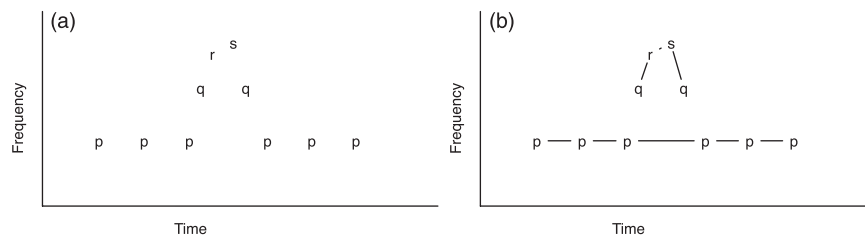

Figure 16. Auditory streams, figure, and ground. A sequence of tones (a) parsed into distinct streams as in (b). One stream or the other may be attentively selected as figure while the other becomes background.

and later parts belong to a single stream. Auditory streams thus are subject to *amodal* completion analogous to what occurs for visual objects partly hidden behind occluding surfaces. Removing segments of a tonal stream and replacing them with broadband white noise in some cases even leads to the illusory experience as of the stream audibly continuing during the presence of the masking noise. In the case depicted in Fig. 17, one seems to hear *both* the masking noise and the continuing stream, despite the absence of a signal corresponding to the original stream (Bregman 28). This auditory illusion as of a continuous stream is analogous to seeing illusory contours when viewing the Kanizsa triangle. The auditory system completes the *temporal* contours of an individual it grasps as continuing while the signal is inaccessible due to masking noise. This involves not just reidentifying an item after a gap, as with seeing a car emerge from a tunnel. Rather, since audible items comprise patterns of qualitative change

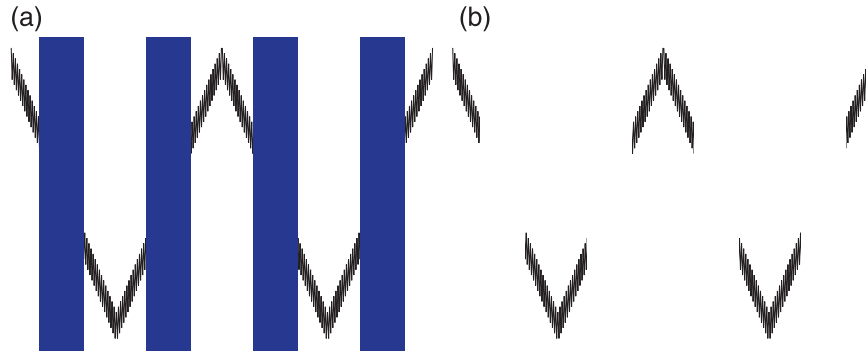(a)   (b)

Fig. 17. Auditory streams and occlusion. (a) An auditory stream varying in pitch over time is heard to continue during the presence of masking noise (represented by the blue bars), even when the signal corresponding to the stream is missing during the presence of masking noise, as in (b).

over time, recognizing the sound stream requires completion through temporal occlusion.

These effects demonstrate that audition distinguishes distinct audible individuals at a time and sequentially integrates adjacent tones into distinct auditory individuals according to principles that parallel processes in the visual perception of objects. Just as visual awareness concerns discrete surfaces that belong to visually integrated objects, auditory awareness concerns discrete sounds assigned either to one integrated stream or another, but not to both. Audition, like vision, assigns discernible individual elements or parts to unified but complex perceptible individuals. Such complex audible individuals are fashioned through the *binding* of audible features, the determination of *edges* and *boundaries*, the *exclusive allocation* of audible components, and the *sequential* (rather than spatial) *integration* of component notes over time. Auditory streams thus comprise a unique variety of perceptual *objects* because they are mereologically complex audible individuals. Their mereological structure and the principles by which they are perceived, however, differ from those of visual objects.

Features of a signal's temporal profile are critical for the individuation at a time and identification over time of auditory streams. Much as spatial features are critical for the individuation and identification of visible objects, auditory objects depend upon time. For instance, coincident onset and attack patterns strongly indicate the presence of a single auditory individual, while different onset times indicate different streams. Temporal gaps or discontinuities frequently mark distinct audible particulars, streams, or sounds, much as spatial discontinuities often indicate different surfaces, while temporally continuous auditory streams are capable of surviving a great deal of qualitative change, much as different parts of a single visible object might bear different features. Nonetheless, temporally extended auditory streams might comprise a sequence of multiple, discrete audible
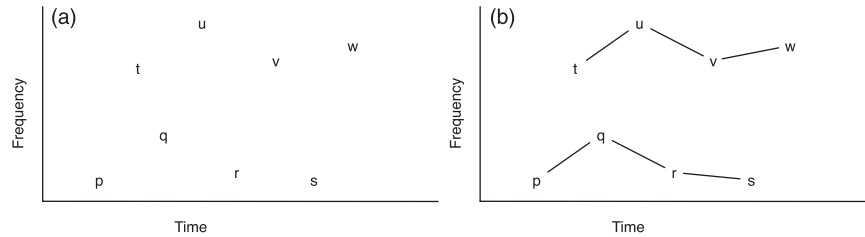
Fig. 18. Simultaneous streams. A sequence of tones (a) heard as comprising distinct but simultaneous streams (b).

individuals, or sounds, just as surfaces may constitute visible objects. Aspects of the contours of change through time, in such cases, are critical to the identification of a single persisting stream.

The space:time::vision:audition analogy, however, is not so neat, since *pitch* also plays an important role in individuating auditory streams (see also, e.g., Handel; Kubovy; Kubovy and Van Valkenburg; Van Valkenburg and Kubovy 2003). Pitch, as previous examples have shown, is important to individuating streams at a time. Since we might hear notes at a single pitch that come from separate speakers as a single audible individual, while different pitches from a single speaker may appear to qualify distinct individuals (see Fig. 14), a difference in pitch may be necessary or even sufficient to hear distinct simultaneous audible individuals, at least across several important kinds of cases. Pitch thus plays a role in individuating auditory individuals at a time similar to that of spatial location in vision.

Pitch also, however, impacts how elements are allocated to auditory streams over time. As in the examples above, when pitch distance is great between successive tones, they are less likely to be integrated into a single continuing stream. In Fig. 18, the pitch distance between successive notes leads subjects to hear two separate streams. Surprisingly, the sequence *p t q u r v s w*, where *p*, *q*, *r*, and *s* are near in pitch and *t*, *u*, *v*, and *w* are near in pitch, sounds like two distinct streams – *p-q-r-s* and *t-u-v-w* – when the pitch distance between the two groups is great (Bregman 17–18).

This suggests a new way to understand the analogy between the role of space in vision and the roles of time and pitch in audition. The roles of time and pitch in fact call attention to two roles of space in visual object perception. Time, I want to suggest, plays a role in audition similar to the role in vision of space – in particular, of spatial extent and boundaries – in determining the *internal* structure and composition of individuals. Just as visual objects appear to fill space and to have spatial parts and boundaries, audible individuals appear to occupy time and to possess temporal parts and boundaries. Pitch, on the other hand, plays a role in audition similar to the role in vision of space – in particular, of spatial location – in determining the structural *relations*, or the *external* structure, among

individuals at a time. Just as different visible individuals have different spatial locations, and just as (all else equal) difference in location suffices for different visual individuals, different auditory individuals have different locations in pitch space, and (all else equal) difference in pitch suffices for difference of auditory individuals (harmonically related tones, which share fundamental frequency, may be an exception). Thus, the space:time::vision:audition analogy must be revised to reflect at least two *different* roles of space in vision. One is its role in determining the structure of a visual object, and the other is its structural role in the visual experience of objects. In audition, time corresponds to the first, while pitch corresponds to the second. Time is the structure internal to auditory objects; pitch is a structure among auditory objects.

What in the world are auditory objects, on the assumption that they are individual auditory streams? To be clear, they are not vision's objects. Auditory objects do not share the spatial and temporal characteristics of ordinary or material objects. They do not seem to have spatial edges, to be opaque like tables and chairs, or to have internal spatial complexity. And auditory objects do not seem wholly to exist at a given moment, as do visual objects. Auditory objects appear to occur, unfold, or take place, and thus occupy time much as visual objects occupy space. In fact, the identities of particular auditory objects may be tied to aspects of patterns of change to audible qualities through time. The sound of the word 'pastoral' differs from the sound of the word 'pasture' precisely because it differs in the pattern of audible qualities it instantiates over time. For this reason, they are perhaps best understood as event-like individuals.

Auditory individuals therefore include what we classify as *sounds*. The perception of auditory streams, however, does not stop at continuous sounds, for multiple discernible sounds interrupted by brief periods of silence might comprise a particular audible stream, such as that of a melody, which can be distinguished from other simultaneous streams. Auditory streams thus may incorporate brief periods of *silence* (on hearing silence, see Sorenson). Auditory objects, construed as auditory streams, may comprise temporally extended sounds or sound sequences that include bits of silence.

What purpose makes the auditory perception of such streams intelligible? I suggest that sounds and streams provide fantastically useful information about their *sources*. Such sources are not just ordinary material objects, understood as such, but what such objects do. Sounds and streams furnish information about the *events* and *happenings* – the collisions, vibrations, and interactions – that commonly make or produce sounds. Auditory objects or streams do not concern the relatively static material objects that exist at a time; they concern the ongoing activities and transactions in which such objects engage. Auditory objects or streams, on this characterization, correspond to audible events that unfold in the material world.[3]

5. Perceptual Objects

I have presented the case that objects are important to understanding the structure and function of both visual and auditory perception. It is time to be explicit about the shared sense in which each counts as a form of object perception.

Vision targets three-dimensional objects of a sort that includes material objects. Shadows, holes, rainbows, and autostereoscopic holograms may belong among visual objects since they share many visible spatio-temporal features with material objects. In particular, they visually appear to exist in their entirety at any given moment – what is present at a moment appears to suffice for being that object.

Audition's objects, however, are not like ordinary material objects. First and foremost, they require time to occur and to unfold. Audition's objects do not strike one as capable of existing entirely at a moment. Auditory objects, or streams, correspond instead to audible happenings, occurrences, or events. Thus, rather than providing immediate awareness of the furniture itself, audition intuitively concerns what the furniture is *doing* – that the couch is being moved, the rocker is rocking, and Junior is kicking the table. Though audible events involve the furniture, audition does not involve perception *as of* the furniture in quite the way that vision and touch do. Audition differs in this from vision and from touch.

What, then, warrants calling audition a form of object perception in any sense stronger than that sound streams are the intentional or proper objects of audition? Does any interesting sense of 'object' remain that captures features of the organization of both vision and audition? A conception of 'perceptual object' that is both broader than 'material object' (and 'visual object') and narrower than 'intentional object' may in fact illuminate the sense in which *all* perception concerns objects.

What makes a certain class of visible or audible individuals perceptual *objects*? The answer, I have suggested, is *composition*. Vision tracks individuals that comprise continuous or coherent collections of spatially bounded parts and surfaces, which correspond most frequently to material objects. Audition tracks individual streams that comprise continuous or coherent collections of temporally bounded tones and sounds, which correspond to interactions or activities in one's environment. The sense in which both vision and audition track individuals that themselves are coherent collections of individuals – surfaces and sounds, respectively – is the sense in which both vision and audition count as forms of object perception. While visual objects are continuous three-dimensional objects, auditory objects are temporally extended sound streams in pitch and physical space. Since they require time to occur or unfold and are not perceptually represented as wholly present at each moment at which they exist, and since they are individuated in terms of pitch and temporal features, sound streams are not objects in the everyday sense. Sound streams, nonetheless, are

mereologically complex individuals tracked by perceptual processes that are strikingly analogous to those deployed in vision. Sound streams therefore are, in a theoretically illuminating sense, perceptual objects.

In what sense, then, is *all* perception about objects, or for the purpose of making objects available for attention, thought, and action? To claim awareness as of ordinary or material objects is central to the organization of experience in each perceptual modality faces serious obstacles, as audition and olfaction demonstrate. Perceptual objects built on the notion of a perceptible individual nonetheless play an important theoretical role in vision, touch, and audition. Such objects are particular items that bear features, persist from one moment to the next, survive certain changes and not others, have boundaries, may comprise parts that are cohesive according to some spatio-temporal or qualitative criterion, and are identified despite occlusion and masking. Perceptual objects, at least for vision, touch, and audition, are mereologically complex but unified individuals.[4]

This represents an important advance over the view that all perceptual objects are ordinary or material objects, as well as over the quietist view that all perception has intentional objects. Vision and audition, perhaps most strikingly, share a predicative structure in which features are assigned to discrete, complex, selectively persisting individuals that are grasped as objective. That vision and audition do not share perceptual objects renders the generalization no less significant. Armstrong is right, at least concerning vision and audition, that perceiving involves attributing features to objects, though vision's objects are a sort different from audition's.

Finally, a caveat. I have been careful not to extend even this generalization to all perception and perceptual experience. It is unclear to me whether gustation and olfaction, for instance, or even kinesthesis, attribute features to individuals, much less to mereologically complex individuals susceptible to figure-ground distinctions and constancies. Extending the verdict without further attention to other perceptual modalities thus is unwarranted.

Short Biography

Casey O'Callaghan's research concerns philosophical issues about perception. In particular, O'Callaghan's work aims to discover what there is to learn about perception by thinking about modalities other than vision. This research recently has focused upon auditory perception and upon the philosophical import of cross-modal perceptual illusions. His book *Sounds: A Philosophical Theory* (Oxford, 2007) presents a philosophical theory of sounds and their perception according to which sounds are events in which objects or interacting bodies disturb a surrounding medium. Recent work proposes that cross-modal perceptual effects and experiences defeat the traditional philosophical and scientific approach which assumes that each modality can be understood exhaustively in isolation from the others. His articles appear in *The Monist*, *Philosophical Issues*, *Oxford Studies in Metaphysics*, *The Cambridge Handbook to Cognitive Science*, and elsewhere. O'Callaghan is Assistant Professor of Philosophy at Rice University, and formerly taught at Bates College in Maine. He received a B.A. in Philosophy from Rutgers University and a Ph.D. in Philosophy from Princeton University.

Notes

★ Correspondence address: Rice University, Department of Philosophy, MS 14, PO Box 1892, Houston, TX 77251-1892, USA.

[1] I do not mean to prejudge the question whether the visually relevant notion of an object itself is a sortal concept (see, e.g., Xu). Rather, such objects need not be understood as belonging to the sortals we deploy in categorizing the *ordinary* objects or paraphernalia we encounter in daily life, such as cups and saucers.

[2] Jim John has suggested to me that hearing a trumpet or two in front of a line of violins might constitute such a case. I am unconvinced that one hears as of a single continuous entity that auditorily seems to spatially complete behind the trumpets.

[3] What about the odd role of pitch? Since audible qualities depend upon distinctive characteristics of ordinary things and their activities, sameness or difference in pitch strongly correlates to one or multiple happenings. When two speakers produce the same note and appear auditorily as a single individual, the experience as of a single sound confirms the unlikely coincidence.

[4] A further shared feature supports this common characterization of vision, touch, and audition. In each modality, perceptual objects are represented to exist in the world independent from one's perceptual experience. Perceptual objects are, in this respect, unlike pains, tickles, and dizziness. Perceptual objects seem distinct from oneself. Several indications mark this sense of objectivity. First, perceptual objects are represented to have distinctive spatio-temporal locations and may appear to continue behind spatial occluders or temporal masking. Second, perceptual constancies demonstrate a perceptual grasp upon the difference between changes to appearances due to observations and those due to changes to perceptual objects themselves. Finally, perceptual objects may be reidentified visually, tactually, or auditorily following a gap in observation. Each modality thus possesses a grasp of the difference between a thing's existence and one's perspective upon it. The perception of visual, tactual, and auditory objects is marked by a capacity to distinguish changes to what is observed from changes to one's experience of what is observed (see Siegel for a useful account of what this amounts to in the case of vision).

Works Cited

Armstrong, D. M. 'In Defense of the Cognitivist Theory of Perception'. *The Harvard Review of Philosophy* 12 (2004): 19–26.

Barlow, H. B. 'Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?' *Perception* 1 (1972): 371–94.

Best, V., F. J. Gallun, A. Ihlefeld, and B. G. Shinn-Cunningham. 'The Influence of Spatial Separation on Divided Listening'. *Journal of the Acoustical Society of America* 120.3 (2006): 1506–16.

Blaser, E., Z. W. Pylyshyn, and A. O. Holcombe. 'Tracking an Object through Feature Space'. *Nature* 408 (2000): 196–9.

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

Clark, A. *A Theory of Sentience*. New York, NY: Oxford UP, 2000.

Cohen, J. 'Objects, Places, and Perception'. *Philosophical Psychology* 17.4 (2004): 471–95.

Feldman, J. and P. D. Tremoulet. 'Individuation of Visual Objects over Time'. *Cognition* 99 (2006): 131–65.

Griffiths, T. D. and J. D. Warren. 'What is an Auditory Object?' *Nature Reviews Neuroscience* 5 (2004): 887–92.

Handel, S. 'Space is to Time as Vision is to Audition: Seductive but Misleading'. *Journal of Experimental Psychology: Human Perception and Performance* 14 (1988): 315–17.

Jackson, F. *Perception: A Representative Theory*. Cambridge: Cambridge UP, 1977.

Kubovy, M. 'Should We Resist the Seductiveness of the Space:Time::Vision:Audition Analogy?' *Journal of Experimental Psychology: Human Perception and Performance* 14.2 (1988): 318–20.

——. and D. Van Valkenburg. 'Auditory and Visual Objects'. *Cognition* 80 (2001): 97–126.

Leslie, A. M. 'The Necessity of Illusion: Perception and Thought in Infancy'. *Thought without Language*. Ed. L. Weiskrantz. Oxford: Clarendon Press, 1988. 185–210.

——, F. Xu, P. D. Tremoulet, and B. J. Scholl. 'Indexing and the Object Concept: Developing "What" and "Where" Systems'. *Trends in Cognitive Sciences* 2.1 (1998): 10–18.

Marr, D. *Vision*. San Francisco, CA: W. H. Freeman, 1982.

Matthen, M. 'The Diversity of Auditory Objects'. *European Review of Philosophy* 7 (2008).

——. *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Oxford UP, 2005.

Nakayama, K., Z. J. He, and S. Shimojo. 'Visual Surface Representation'. *An Invitation to Cognitive Science, Vol. 2, Visual Cognition*. Eds. S. M. Kosslyn and D. N. Osherson. 2nd ed. Cambridge, MA: MIT Press, 1995. 1–70.

O'Callaghan, C. *Sounds: A Philosophical Theory*. Oxford: Oxford UP, 2007.

Pasnau, R. 'What is Sound?' *Philosophical Quarterly* 49 (1999): 309–24.

Pylyshyn, Z. W. *Things and Places: How the Mind Connects With the World*. Jean Nicod Lecture Series. Cambridge, MA: MIT Press, 2007.

——. 'Visual Indexes, Preconceptual Objects, and Situated Vision'. *Cognition* 80 (2001): 127–58.

Scholl, B. and Z. Pylyshyn. 'Tracking Multiple Items through Occlusion: Clues to Visual Objecthood'. *Cognitive Psychology* 38 (1999): 259–90.

Scholl, B. J. 'Objects and Attention: The State of the Art'. *Cognition* 80 (2001): 1–46.

Sekuler, R., A. B. Sekuler, and R. Lau. 'Sound Alters Visual Motion Perception'. *Nature* 285 (1997): 308.

Shoemaker, S. 'Qualities and Qualia: What's in the Mind'. *Philosophy and Phenomenological Research* 50 (1990): 109–31.

Siegel, S. 'Subject and Object in the Contents of Visual Experience'. *Philosophical Review* 115.3 (2006): 355–88.

Sorenson, R. *Seeing Dark Things*. New York, NY: Oxford UP, 2007.

Spelke, E. S. 'Principles of Object Perception'. *Cognitive Science* 14 (1990): 29–56.

——. 'Where Perceiving Ends and Cognition Begins: The Apprehension of Objects in Infancy'. *Perceptual Development in Infancy*. Ed. A. Yonas. Hillsdale, NJ: Lawrence Erlbaum, 1988. 197–234.

Van Valkenburg, D. and M. Kubovy. 'In Defense of the Theory of Indispensible Attributes'. *Cognition* 87 (2003): 225–33.

Wertheimer, M. 'Experimental Studies on the Seeing of Motion'. *Classics in Psychology*. Ed. T. Shipley. New York, NY: Philosophical Library, 1912/1961.

Wilson, J. A. and J. O. Robinson. 'The Impossibly Twisted Pulfrich Pendulum'. *Perception* 15 (1986): 503–4.

Xu, F. 'From Lot's Wife to a pillar of Salt: Evidence for *Physical Object* as a Sortal Concept'. *Mind and Language* 12 (1997): 365–92.

Supplementary Material

The following supplementary material is available for this article online:

Video Clip S1. Motion cues to objecthood. <http://www.blackwell-compass.com/home_video#phco1>. See also Fig. 8.

Video Clip S2. Streaming and bouncing. <http://www.blackwell-compass.com/home_video#phco2>. See also Fig. 10.

Video Clip S3. Multiple object tracking. <http://www.blackwell-compass.com/home_video#phco3>. See also Fig. 11.

This material is available as part of the online article from http://www.blackwell–synergy.com