Cognition 170 (2018) 88–94

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents

Hauke S. Meyerhoff[a,*], Brian J. Scholl[b]

[a] Leibniz-Institut für Wissenmedien, Tübingen, Germany
[b] Yale University, New Haven, CT, United States

ARTICLE INFO

Keywords:
Audiovisual integration
Bouncing/streaming
Causal perception
Object persistence

ABSTRACT

A central task for vision is to identify objects as the same persisting individuals over time and motion. The need for such processing is made especially clear in ambiguous situations such as the bouncing/streaming display: two discs move toward each other, superimpose, and then continue along their trajectories. Did the discs stream past each other, or bounce off each other? When people are likely to perceive streaming, playing a brief tone at the moment of overlap can readily cause them to see bouncing instead. Recent research has attributed this effect to decisional (rather than perceptual) processes by showing that auditory tones alter response biases but not the underlying sensitivity for detecting objective bounces. Here we explore the nature of this phenomenon using 'illusory causal crescents': when people perceive bouncing (or causal 'launching'), they also perceive the second disc to begin moving before being fully overlapped with the first disc (i.e. leaving an uncovered crescent). Here we demonstrate that merely playing a sound coincident to the moment of overlap can also reliably induce the perception of such illusory crescents. Moreover, this effect is due to the coincidence of the tone, per se, since the effect disappears when the tone is embedded in a larger regular tone sequence. Because observers never have to explicitly categorize their percept (e.g. as streaming)—and because the effect involves a subtle quantitative influence on another clearly visual property (i.e. the crescent's width)—we conclude that this audiovisual influence on the perception of identity over time reflects perceptual processing rather than higher-level decisions.

1. Introduction

Although the light that enters our eyes is continuous, the world that we consciously perceive is often discrete, consisting of arrangements of individual objects. Object individuation does not end with segmentation, however: beyond identifying some feature cluster as an object, we must also identify it as the *same* object over time and motion—and sometimes over featural changes and interruptions such as occlusion. As a result, a great deal of visual processing is involved in computing so-called *object persistence* over time (for a review, see Scholl, 2007).

1.1. Bouncing vs. streaming

The perception of object persistence is required during nearly every waking moment of our lives: without it, our visual experience would likely be incoherent, with objects popping into and out of existence haphazardly. But the need for such processing is especially apparent in ambiguous situations wherein there are two or more salient possible options for 'which went where' over time. A perfect example of this situation is the bouncing/streaming display, which grew out of the Gestalt tradition of perception research (Metzger, 1934; for a review see Wagemans et al., 2012). In this display, two objects move toward each other (either in a single plane, or on diagonal trajectories) until they are superimposed, at which point two objects continue along those same trajectories until they are fully separated. This display can be perceived as either two objects bouncing off each other (following a sort of central collision, after which each object suddenly reverses its trajectory) or as two objects streaming past each other (without interacting at all, despite meeting in the middle).[1]

Perhaps because this ambiguity can be so striking (no pun intended)—and because the two percepts seem so categorically distinct—this sort of display has often been used to test various manipulations related to object persistence. (In addition, such displays don't require the same sorts of tight spatiotemporal constraints as do related displays such as the Ternus configuration in studies of apparent motion correspondence; e.g. Kramer & Yantis, 1997; Ternus, 1926.) For example, one can explore whether the perception of bouncing vs. streaming is influenced by the congruence of surface features such as

* Corresponding author at: Leibniz-Institut für Wissenmedien, Schleichstr. 6, 72076 Tübingen, Germany.
E-mail address: h.meyerhoff@iwm-kmrc.de (H.S. Meyerhoff).

[1] This sort of display was first studied by Metzger (1934), Michotte (1946/1963), Experiments 24 and 97), and Julesz (1959, cited in Julesz, 1995, p. 50).

http://dx.doi.org/10.1016/j.cognition.2017.08.007
Received 20 January 2017; Received in revised form 14 August 2017; Accepted 17 August 2017
0010-0277/

color and shape (e.g. Caplovitz, Shapiro, & Stroud, 2011; Feldman & Tremoulet, 2006; Shapiro, Caplovitz, & Dixon, 2014), by spatial alignment (e.g. Kawabe & Miura, 2006), by spatial offsets in the motion paths (e.g. Grove, Robertson, & Harris, 2016; Grove & Sakurai, 2009), by variations in object speed (e.g. Sekuler & Sekuler, 1999; Zeljko & Grove, 2017a), by the presence of central occlusion (e.g. Remijn & Ito, 2007), or by the creation and maintenance of 'object files' (e.g. Mitroff, Scholl, & Wynn, 2005).

1.2. "Boom!": An audiovisual interaction?

When faced with a bistable display such as bouncing/streaming, the visual system can draw on many potential cues to disambiguate it—including cues from other modalities. For example, when people are likely to perceive streaming, playing a brief 'click' at the moment of overlap can readily cause them to see bouncing instead (Sekuler, Sekuler, & Lau, 1997; see also Grassi & Casco, 2010; Sanabria, Correa, Lupiáñez, & Spence, 2004; Zhou, Wong, & Sekuler, 2007). In this case, the temporally-aligned 'click' can be readily explained by appeal to a 'bounce', whereas it would otherwise be an unexplained coincidence if it just happened to occur at the moment of overlap during streaming.

This underlying 'logic of perception' (cf. Rock, 1983) can be appreciated by exploring the temporal synchrony of the click with the moment of full visual overlap: the effect is strongest with perfect synchrony, and by the time there is more than a couple hundred milliseconds of asynchrony, the effect vanishes (Sekuler et al., 1997). (Similar effects occur for coincident visual flashes; e.g. Watanabe & Shimojo, 1998.) And the influence of this temporal coincidence is so powerful that it can yield bouncing percepts even with spatial offsets between the two objects that would otherwise render bouncing unlikely (Grove & Sakurai, 2009; but see Grassi & Casco, 2009). Moreover, removing the 'coincidence' eliminates the effect: if the sound occurs as part of a temporal sequence of identical tones that extends both before and after the moment of overlap, then the temporally-aligned tone (now one among many identical tones) no longer leads to perceived bouncing, as it would in isolation (or with a tone that deviates from the other tones in the temporal sequence; Watanabe & Shimojo, 2001; see also Kawachi & Gyoba, 2006).

1.3. Seeing or thinking?

Although we have been describing the audiovisual influence on bouncing/streaming as an effect on what we see (cf. Dufour, Touzalin, Moessinger, Brochard, & Després, 2008), others have suggested that it is an effect on (merely) our higher-level *decisions* about what must have happened in the display. In particular, a recent study employed signal detection theory to distinguish changes in sensitivity from changes in response biases in the context of bouncing vs. streaming—where sensitivity is associated with perceptual processes, and biases are associated with "decisional processes" (which are to be contrasted with perception.) These researchers (Grove, Ashton, Kawachi, & Sakurai, 2012) disambiguated the bouncing/streaming display by introducing objects with different visual features; despite this disambiguation, coincidental sounds still led to increased reports of bouncing—but this effect reflected a change in response bias rather than a change in sensitivity, leading the authors to conclude that the effect occurs at a "decisional level" rather than at a "perceptual level" (p. 2). And such interpretations are consistent with a broader theoretical perspective wherein such causal events are treated as the result of higher-level interpretations rather than seeing, per se (Rips, 2011).

This conclusion strikes us as rather unlikely for several reasons. *First*, the logic of signal detection theory simply does not apply in this context, when it comes to distinguishing perception vs. cognition. The influence of the temporally aligned sound on perceived bouncing vs. streaming is precisely an influence on our (inherently subjective) *experience*, not on any sort of objective detection. So in this context, a criterion shift (as measured by signal detection) can reflect a change in perception just as much as a sensitivity shift (see Grassi & Casco, 2012; see also Witt, Taylor, Sugovic, & Wixted, 2015). (And of course, this logic also applies to many, if not most, other visual illusions: in so many cases, observers are not sensitive to some objective feature of the image [such as the true lengths of lines], and are instead responding in a 'biased' way [e.g. being subject to the Muller-Lyer illusion]. Such 'biases' do not imply that these illusions are only higher-level decisions.) And critically, this argument is independent of the exact nature of the signal detection task (see Witt et al., 2015; Zeljko & Grove, 2017b).

Second, because the auditory-induced bouncing effect is an effect on what we see, observers can appreciate the effect first-hand, without relying on any analyses or arguments (see Firestone & Scholl, 2016). And indeed, we suspect that one of the reasons why this phenomenon has attracted so much interest over the years is precisely because it works so well as a vivid *demonstration*—in which we can appreciate the change in our perceptual experience, and not just in our post hoc decisions. *Third*, there is plenty of other evidence that the perception of bouncing vs. streaming (in the context of perceived 'launching' vs. 'passing') reflects truly perceptual processing—since (for example) it can yield retinotopically specific visual adaptation (Rolfs, Dambacher, & Cavanagh, 2013; see also Kominsky & Scholl, submitted for publication).

Nevertheless, we must admit that these arguments each may leave something to be desired. The first argument does not imply that the effect *is* perceptual, but merely that it *could* be. The second argument appeals only to direct subjective experience, and so may seem more or less compelling depending on one's own phenomenology. And the third argument does not directly address the audiovisual effect in question: just because some other (purely visual) influences on bouncing vs. streaming may be clearly perceptual does not entail that all (much less these particular crossmodal) effects must be.

1.4. The current study: From sound to crescents?

How could we test more directly whether auditory influences on bouncing vs. streaming reflect perceptual processing or merely "decisional" effects? One promising strategy would be to explore whether this effect can also influence other independent aspects of our perceptual experience. Here we do so, by testing whether a temporally aligned sound can not only drive bouncing percepts, but can also change the perceived spatial relationships among the objects involved in the putative event. In fact, previous data has revealed that observers may misperceive some spatial relations between the moving discs when the overlapping event is coincident with a brief tone (Kawachi, 2016; see also Grassi & Casco, 2012). However, we are not aware of any previous work that used such manipulations to directly investigate whether the tone-induced influences on bouncing/streaming displays stem from truly perceptual or decisional processes. To do so here, we exploit the phenomenon of 'illusory causal crescents' (Scholl & Nakayama, 2004).

An analogue of bouncing vs. streaming occurs in the perception of causal launching (Michotte, 1946/1963; for a brief review see Wagemans, Van Lier, & Scholl, 2006): one disc (A) moves toward a second stationary disc (B) until they fully overlap, at which point one of the discs continues moving and the other remains still. This display is again ambiguous: it can be perceived as A causally 'launching' B (so that at the moment of overlap, A stops and B starts moving), or as A 'passing' over (or through) an always-stationary B. Other contextual events—such as an unambiguous launching event, wherein B begins moving as soon as A is adjacent to it—can force observers to perceive launching in the full-overlap event, however, as long as the two events are temporally aligned (Scholl & Nakayama, 2002; see also Choi & Scholl, 2004). (This effect has also been recently demonstrated in non-human primates; Matsuno & Tomonaga, 2017.)

If a full-overlap event is seen as causal launching, however, then the two discs could not have actually overlapped completely. From the