BRILL

Multisensory Research 29 (2016) 453–464

MULTISENSORY RESEARCH

brill.com/msr

Disambiguating the Stream/Bounce Illusion With Inference

Philip M. Grove [1,*], **Caitlin Robertson** [1] **and Laurence R. Harris** [2]

[1] School of Psychology, The University of Queensland, Australia

[2] Centre for Vision Research and Department of Psychology, York University, Canada

Received 28 June 2015; accepted 7 November 2015

Abstract

The ‘stream/bounce’ illusion refers to the perception of an ambiguous visual display in which two discs approach each other on a collision course. The display can be seen as two discs streaming through each other or bouncing off each other. Which perception dominates, may be influenced by a brief transient, usually a sound, presented around the time of simulated contact. Several theories have been proposed to account for the switching in dominance based on sensory processing, attention and cognitive inference, but a universally applicable, parsimonious explanation has not emerged. We hypothesized that only cognitive inference would be influenced by the perceptual history of the display. We rendered the display technically unambiguous by vertically offsetting the targets’ trajectories and manipulated their history by allowing the objects to switch from one trajectory to the other up to four times before the potential collision point. As the number of switches increased, the number of ‘bounce’ responses also increased. These observations show that expectancy is a critical factor in determining whether a bounce or streaming is perceived and may form the basis for a universal explanation of instances of the stream/bounce illusion.

Keywords

Stream/bounce, audio/visual integration, multisensory perception, motion, inference

1. Introduction

A simple but tractable audio-visual phenomenon, called the ‘stream/bounce’ effect, has been used extensively to study audio-visual perception (e.g., Fujisaki *et al.*, 2004; Sekuler *et al.*, 1997; Zhou *et al.*, 2007). Typically, two identical discs appear on either side of a display and move towards one another along collinear horizontal trajectories. Critically, the two discs completely su-

* To whom correspondence should be addressed. E-mail: p.grove@psy.uq.edu.au

DOI:10.1163/22134808-00002524

perimpose at the midpoint of their trajectory (hereafter referred to as the *point of coincidence*, POC) and continue to the other side of the display to take up the position of the other disc. When viewing visual-only sequences, observers typically report that the discs 'stream' past one another on approximately 70% of trials (Sekuler *et al.*, 1997; Watanabe and Shimojo, 2001a). However, for conditions in which a transient such as a brief sound is presented at or near the POC, the bias reverses such that observers report 'bouncing' (i.e., the discs reverse their motion after coinciding) on approximately 70% of trials (Fujisaki *et al.*, 2004; Grove and Sakurai, 2009; Sekuler *et al.*, 1997). This is a robust phenomenon that has been replicated in several contexts. For example, the stream/bounce effect generalizes to motion-defined (Burns and Zanker, 2000) and cyclopean (Grove *et al.*, 2012a) targets and also manifests *via* indirect measures such as reaction time to detect objective streaming and bouncing events (Sanabria *et al.*, 2004, 2007). Three classes of hypothesis based on sensory processing, attention, or inference, have been proposed to account for the transient induced switching of the dominant resolution of stream/bounce displays but none have provided a parsimonious explanation.

1.1. The Sensory Processing Hypothesis

The sensory processing hypothesis is a bottom-up hypothesis that postulates that the integration of the discs' motion is somehow altered at the moment of coincidence such that the information available to the brain is different in a way that encourages one or other interpretation. For example, Bertenthal *et al.* (1993) attributed the bias towards streaming perception in transient-free displays to 'temporal recruitment' (Anstis and Ramachandran, 1987), a putative sensory process in which motion filters tuned to a given velocity interact in an excitatory manner for that velocity but inhibit one another at other velocities. These processes are thought to bias the network to detect motion in the same direction and speed as previously stimulated. Kawabe and Miura (2006) reported that oriented targets (Gabor patches) generated more streaming responses when they were collinear with their motion trajectory than when their orientation deviated from collinearity. Kawabe and Miura argued, in agreement with Bertenthal *et al.* that target orientation strengthened the spatiotemporal integration of local motion signals. Sekuler and Sekuler (1999) examined Bertenthal *et al.*'s temporal recruitment hypothesis in more detail but failed to find support for it in all conditions, leading them to pose an inference hypothesis, described below. Another difficulty for a sensory processing account is that observers' ability to discriminate between objective streaming and objective bouncing events is unaffected by sound. That is, their sensitivity (d') is the same whether or not a sound is presented at the POC (Grove *et al.*, 2012b). Signal detection analysis does, however, reveal significant changes in response criterion depending on the presence or absence of sound at the POC.

Response criterion is significantly more liberal when the sound is presented at coincidence compared to when it is absent, consistent with a response bias underlying the effect rather than a change in sensitivity (Grove *et al.*, 2012b).

1.2. The Attention Hypothesis

The attention hypothesis posits that the transient momentarily disrupts attention resources deployed to track the moving objects which disrupts in some unspecified way the establishment of the correspondence of the objects' path (Kawabe and Miura, 2006; Kawachi and Gyoba, 2013). Watanabe and Shimojo (1998) proposed that temporal recruitment requires attention in order to be maintained. They suggested that distracting attention from the moving targets when they are close to the POC disrupts the temporal recruitment process resulting in the perception of bouncing. However, Grassi and Casco (2009) showed that equally salient sound profiles could generate different interpretations. Furthermore, Grassi and Casco (2010) demonstrated that sounds congruent with a collision, such as the sound of billiard balls colliding, collect more bounce responses than sounds that are not congruent with a collision, such as a water drop or a firecracker sound. The authors argued that attention plays a minor role based on their observations that reaction times to all three types of sounds were similar, suggesting that the three sounds were equivalent in terms of attracting attention. Another interesting finding from this study was that the billiard-ball-collision sound dominated only when the sounds were presented 200 ms prior to coincidence and not when the sounds were presented at coincidence. This implies that the perceptual system requires time to infer a link between the auditory stimulus and the visual event. This idea is elaborated below.

1.3. The Inference Hypothesis

A hypothesis that involves decisional processes is reminiscent of Helmholtz's 'unconscious inference', which he postulated to solve the inverse optics problem: how can the brain decide which of an infinite range of possible stimuli created a particular retinal image? Watanabe and Shimojo (2001a, b) and Sekuler and Sekuler (1999) argue that audiovisual perception must solve the inverse physics problem. In so doing, the perceptual system combines the available sensory information to infer the nature of the external event in a probabilistic way. In the context of the audio-visual stream bounce effect, the perceptual inference invoked to solve trials with a sound at the POC would be based on an auditory input — an impact sound at or temporally close to the POC, and a visual input — the sight of two objects contacting one another. Thus, the motion sequence paired with a sharp sound at or near the point of coincidence tips the balance of likelihood towards a 'bounce' interpretation.

A motion sequence viewed without a sound tips the balance towards a streaming interpretation.

Perturbing either the visual or auditory input can affect the probability of which interpretation is chosen. Considering perturbations of the sound input first, in their experiments, which lead them to propose the inverse-physics account, Watanabe and Shimojo (2001b) employed an ambiguous visual input consisting of identical discs on a collinear trajectory completely overlapping at the POC. Their critical auditory input was a brief sound at coincidence whose effectiveness in inducing bouncing was reduced (but not eliminated) when similar flanking sounds were presented prior to or after coincidence.

Considering perturbations to the visual input, the trajectories of the moving discs may deviate from being perfectly collinear and therefore less compatible with a bouncing resolution, though such a deviation does not preclude the interpretation that the discs collided when combined with an impact-like sound at the POC (Grove and Sakurai, 2009; Grove *et al.*, 2012b). Indeed, Grove *et al.* demonstrated that the probability of a bounce interpretation is significantly higher in the presence of a sharp collision sound than when no sound is present and this difference persists even when the targets differ in appearance by nine JNDs (Just Noticeable Differences).

The inverse physics inferential account implies that the stream/bounce display is resolved at a rather high level of perceptual processing involving cognitive processes, expectations, interpretation, and inference based on mappings of auditory and visual inputs onto events in the world. Therefore, we hypothesised that whether the display is interpreted as streaming or bouncing should be influenced by the perceptual history of the display during which expectations should be built up that influence the final decision. Any number of events could occur as the discs approach the POC. For example, the discs could momentarily pause along the way, change colour, or momentarily disappear. We decided to adapt the displays used by Grove and Sakurai (2009) to manipulate the motion history of the discs prior to the POC. Grove and Sakurai (2009) established that the sound-induced bias towards bouncing responses persists even when a vertical offset is introduced between the trajectories of the moving discs, though the bias towards bouncing reduces as the vertical offset is increased. We took advantage of this as an intuitive way to vary the perceived probability of an impact by having the discs switch tracks as they moved towards each other. By showing one or more switches before the POC we could modulate the expectation of impact. This could only have an effect on inverse physics computations (what is the most likely external event creating the perception?) by priming the putative inferential process. Both ‘attention’ and ‘sensory’ hypotheses predict no effect of the prior history of the moving discs. As well as varying the number of track switches, we also varied the magnitude of the vertical offset between the target trajectories with the expectation that fewer bounces would be reported for a larger vertical offset than for a smaller one. Our predictions are (1) the introduction of a constant vertical offset between the target trajectories will attenuate or eliminate the sound-induced bounce effect; (2) introducing trajectory switches will increase the probability of bounce responses as the number of switches increases. Either or both of these outcomes would represent strong support for the inference hypothesis.

2. Methods

2.1. Participants

Twenty-eight volunteers (25 female) participated. All participants had normal or corrected-to-normal visual acuity and, by self-report, had normal hearing. These experiments were approved by the ethics board of the University of Queensland and conformed to the Treaty of Helsinki.

2.2. Apparatus and Stimuli

Stimuli were generated and scripted on a Macintosh computer (Operating system 9.2.2), running Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) for Matlab (Version 5.1). Stimuli were presented on a 19-inch Samsung Sync Master 1100P Plus CRT. Screen resolution was set to 1024×768 pixels and a refresh rate of 100 Hz. One pixel subtended 1.6 min arc at the viewing distance of approximately 100 cm. Participants' heads were unrestrained during the experiment.

Visual motion sequences consisted of 73 frames, each lasting 30 ms and depicted two identical black discs (diameter 32 min arc, i.e., 20 pixels) moving across the display on a white background. The discs occupied identical x-coordinates on the display in frame 37 of the motion sequence, the POC. A fixation cross was present in all trials, positioned 86.4 min arc below the centre of the display. The discs were displayed initially separated horizontally by 11.52°. The horizontal movement of the discs was at a constant velocity of 5.3°/s. This is shown in Fig. 1.

Sounds were delivered *via* two Audio Excel MS-120 speakers positioned on either side of the stimulus display. Sounds, when present, were presented at the beginning of frame 37 (the POC frame) of the motion sequence. The transient sound was 800 Hz, duration 8 ms (no ramp modulation at onset), and 66 dB SPL at the participant's ear. The ambient sound level in the room was 46 dB SPL.

We generated motion sequences with three different vertical offsets, zero offset or collinear, 6.4 min arc (4 pixels), and 19.2 min arc (12 pixels) offset. The non-zero offsets were chosen based on pilot experiments and on previous work (Grove and Sakurai, 2009) to ensure that the vertical offsets were

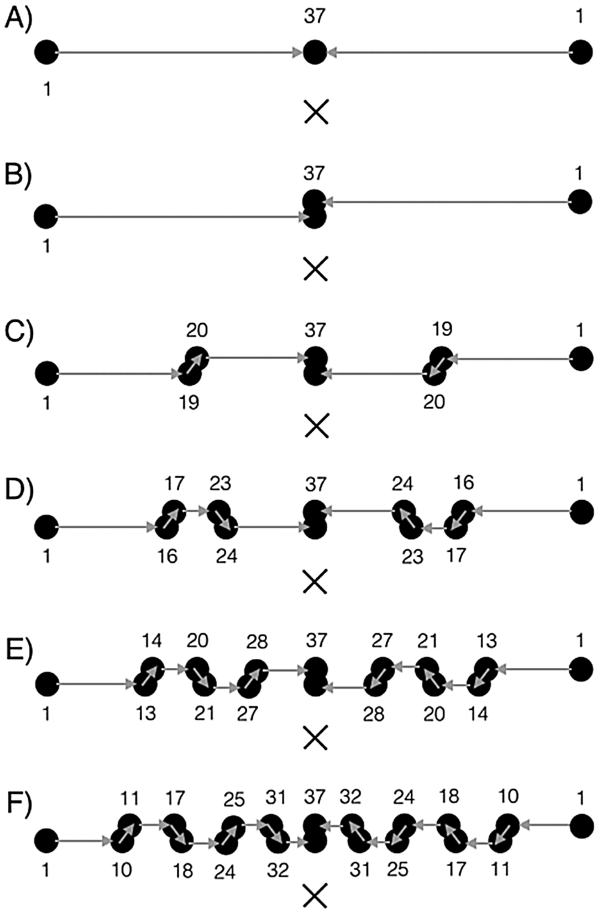

Figure 1. Scale illustration of critical epochs of the motion sequence leading up to the POC for the six different motion sequences employed (A–F). Dot size, horizontal step-size between frames, and vertical offsets are all proportional to the actual stimulus display (19.2′ offset illustrated here). Numbers above the dots indicate the frame number corresponding to that position. Grey arrows represent the path of the discs and were not present in the experimental displays. The cross represents the fixation marker. In (A) the conventional stream/bounce visual display with collinear trajectories; (B) zero switches (constant vertical offset); (C) targets switch once before coincidence (frames 19 and 20); (D) targets switch twice before coincidence (first in frames 16 and 17; then in frames 23 and 24); (E) targets switch three times before coincidence (first at frames 13 and 14, second at frames 20 and 21 and third at frames 27 and 28; (F) targets switch four times before coincidence (first at frames 10 and 11, second at frames 17 and 18, third at frames 24 and 25, fourth at frames 31 and 32). Note that targets remained on a straight trajectory (no switches) from frame 37 to frame 73 (see text). Example stimuli are available online as Supplementary Material.

well above threshold but not too large to eliminate sound-induced bounces. The collinear (zero offset) trajectory traced along the centre of the display (Fig. 1A). In the non-zero vertical offsets, the left disc was displaced upward from the collinear trajectory by half the offset and the right disc was displaced downward by half that value (Fig. 1B–F). On half the displacement trials the left disc was displaced above the centre line and the right disc below. The opposite was the case on the other half of the trials. For each vertical offset, we generated five 'switch' conditions in which the discs switched trajectories 0, 1, 2, 3, or 4 times.

2.3. Procedure

Participants sat 100 cm in front of the stimulus display. Chair height was adjusted so that their eyes were approximately at the same height as the centre of the display. Participants initiated the motion sequences by pressing the space bar on a computer keyboard. After viewing the entire motion sequence, participants indicated if they perceived the discs bounce (i.e., reverse their motion after coinciding at the centre of the display) by pressing the B key. If the discs appeared to stream past one another (i.e., their motion did not reverse after coinciding), participants pressed the S key. Non-zero (6.4 and 19.2 min arc) vertical offset conditions were blocked, with block order counterbalanced across observers. Each block consisted of a collinear condition and five switching conditions. All visual trajectory and sound conditions were randomised within each block. For each block, participants completed 20 trials in the collinear condition with no sound and 20 trials with a sound at the POC; 20 trials in the vertical-offset condition with zero switches and no sound, and 20 trials with a sound at the POC; likewise for the vertical offset condition with 1, 2, 3, and 4 switches. Therefore, there were 120 sound-absent trials, 120 sound-present trials presented in a random order within each block for a total of 240 trials. Participants completed two blocks (6.4 and 19.2 arc min vertical offsets) for a grand total of 480 trials.

3. Results

The mean percentage of bounce responses in each condition are plotted in Fig. 2 for the small (Fig. 2A) and large (Fig. 2B) trajectory separation, as a function of the number of reversals preceding the POC.

All statistical tests were evaluated using the Greenhouse Geisser corrected degrees of freedom though the uncorrected degrees of freedom are reported. A 2 (vertical offset: 6.4; 19.2 min arc) $\times$ 2 (sound: present; absent) $\times$ 6 (number of switches: collinear, 0, 1, 2, 3, 4) within-participants ANOVA revealed a significant main effect for vertical offset, $F(1, 27) = 13.03$, $p = 0.001$, $\eta_p^2 = 0.33$, with more bounces reported overall in the 6.4 min arc offset

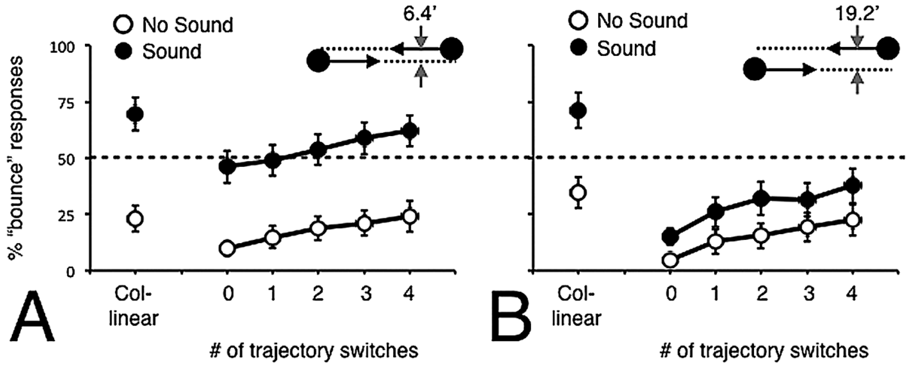

Figure 2. Group mean percentage of bounce responses ($n = 28$) for no-sound (open circles) and with-sound (filled circles) conditions for 6.4′ (A) and 19.2′ (B) vertical offset between trajectories as a function of the number of switches. Leftmost data points indicate the percentage of bounce responses for the collinear stream/bounce. Error bars indicate ±1 SEM.

condition than the 19.2 min arc offset condition; a main effect for sound $F(1,27) = 34.4$, $p < 0.001$, $\eta_p^2 = 0.56$, with more bounces reported overall in the sound present conditions than the sound absent conditions; and a main effect for number of switches $F(5,135) = 6.6$, $p = 0.009$, $\eta_p^2 = 0.2$, with more bounces reported as more trajectory switches occurred prior to coincidence. There was a significant interaction between vertical offset and sound conditions $F(1,27) = 15.2$, $p = 0.001$, $\eta_p^2 = 0.36$, indicating that sound promoted bouncing responses to a greater extent in the 6.4 min arc vertical offset condition than in the 19.2 min arc offset condition; a significant interaction between vertical offset and number of switches $F(5,135) = 8.7$, $p = 0.001$, $\eta_p^2 = 0.24$, indicating that the increase in bouncing reports as a function of trajectory switches was greater when the vertical offset was small (6.4 min arc) than when it was large (19.2 min arc); a significant interaction between sound and number of switches $F(5,135) = 6.5$, $p = 0.004$, $\eta_p^2 = 0.2$, indicating that the increase in bouncing reports as a function of trajectory switches was greater with a sound at coincidence than with no sound.

Inspection of Fig. 2 shows that the collinear conditions (left hand data points) show the typical response pattern with streaming reported on the majority of trials (∼70–75%) in the no sound condition and bouncing reported on ∼75% of trials in the sound condition. In the vertical offset conditions, however, nearly all the data points are below the 50% line, indicating that streaming was the dominant response in most of these conditions. This is expected because the vertical offset partially disambiguates the motion sequence and biases the interpretation towards streaming. The observation that a sound results in significant increases in bouncing responses above the no sound condition across all trajectory switch conditions is evidence of the organising strength of a sound on the resolution of visual motion sequences.

4. Discussion

The present study replicates and extends previous findings on the stream/bounce effect. We confirmed the earliest reports that sound induces a bias towards perceiving bouncing (Sekuler *et al.*, 1997). We also confirmed that the sound-induced perception of bouncing was reduced by introducing a vertical offset between the trajectories, consistent with Grove and Sakurai (2009) and Grove *et al.* (2012a). These observations can be explained by all three potential mechanisms: attention, sensory processing and inference. However, our novel finding is that the reduction of perceived bouncing due to vertical trajectory offsets was modulated by introducing events during the pre-coincidence trajectory which suggests that an event consistent with a motion reversal is possible at the point of coincidence. The modulating effect of the trajectory switches prior to coincidence can only be explained by an inferential process solving the inverse physics problem associated with these audio-visual events. This is the most parsimonious explanation of the inherent tendency to perceive streaming when the trajectories are displaced vertically.

4.1. Perceptual Changes not Involved

Previous reports have claimed that transient-induced motion reversals may be driven at least in part by perceptual processes. Grassi and Casco (2012) used signal detection theory to characterise observers' sensitivity (d') and response criterion when discriminating between partially (overlap 60% and 80%) and completely overlapping discs. They found that observers' sensitivities were lower in sound conditions than in no-sound conditions, in addition to a more liberal response criterion. In order to link these measurements to the stream/bounce effect, Grassi and Casco conducted a second experiment in which they confirmed the conventional promotion of bouncing in the presence of a sound compared to no-sound conditions in the same displays correlated with the changes in sensitivity. The authors inferred an effect of sound on both sensitivity and response criterion contributed to biasing responses toward streaming or bouncing in these displays. However, their sensitivity and criterion measures were for perceived overlap of the discs and did not address the resolution of the motion sequences. These results conflict with those reported by Grove *et al.* (2012b) who found no change in sensitivity but significant changes in criterion as a function of the presence or absence of sound at the POC when discriminating between objective streaming and bouncing events.

A critical difference between Grassi and Casco (2012) and Grove *et al.* (2012a) is that Grove *et al.*'s participants discriminated between objectively bouncing and objectively streaming motion sequences after viewing the entire motion sequence while Grassi and Casco's participants were required to make discriminations about the objective overlap of the discs at the POC. When participants are required to make judgments about associated events, such as the overlap of the discs, at the exact moment they are occurring, sensitivity was affected by the presence of sound (Grassi and Casco, 2012). This may be correlated with streaming or bounce perception but does not determine whether one or the other perception is reported. However, when observers report their perception of an entire sequence they are more likely to show criterion shifts (Grove *et al.*, 2012b) as a function of sound but no change in sensitivity. In the present study, the information was distributed over the entire motion sequence and participants indicated whether the targets appeared to stream or bounce. Thus, the information provided and the perceptual judgements were most consistent with an inferential strategy.

4.2. The Inference Model

Our data support the 'parsimonious perceptual inference account' originally proposed by Sekuler and Sekuler (1999) and Watanabe and Shimojo (2001a, b) to account for the perceptual resolution of motion sequences as either streaming or bouncing. We show that the perceptual system integrates information from the 37 frames (1110 ms) before the POC — quite outside the conventional 'temporal integration window' for these displays (Fujisaki *et al.*, 2004) — and that this integration affects what is perceived at the POC. Switches between 810 and 570 ms prior to the POC had a pronounced effect on perceived bouncing, whereas previous reports consistently show that a sound is ineffective in promoting perceived bouncing if presented more than 350 ms prior to the POC (e.g., Fujisaki *et al.*, 2004; Kawachi *et al.*, 2014). Visual transients need to be even closer to the event to be effective (Shimojo and Shams, 2001). We therefore interpret our data as suggesting that our subjects made an inference about the external event based on the discs' previous history well outside the conventional temporal integration window.

The salience of our trajectory switches increased as their number increased. This is in contrast to Watanabe and Shimojo (2001b) who suggested that multiple auditory events around the POC actually *reduced* the salience of the critical sound presented at the POC. Here we have shown that visual events (trajectory switching) *accumulate influence* with increasing number such that the more switches that occur prior to coincidence, the more likely the observer is to expect an abrupt change in trajectory and the higher the probability that 'bouncing' is reported.

We have reported data that support the inverse physics account, an account that involves a rather high level of perceptual processing involving cognitive processes. Some future lines of investigation might include: (1) generalising our findings to displays in which the discs' motions are random and independent of one another rather than being constrained to two trajectories with structured switches as was employed here. Our prediction would be that more frequent abrupt transitions prior to coincidence would increase the likelihood of bounces being reported, particularly on sound trials; (2) systematically investigating the effect of expectations built up over several trials on the resolution of stream/bounce displays. For example, stream/bounce sequences with visually distinguishable discs (e.g., a white one and a black one against a grey background) could be generated in which the targets either objectively streamed or objectively bounced. The two types of sequences would be presented multiple times but with different probabilities of each occurring within an experimental block thus manipulating the participant's expectations. Test stimuli, consisting of conventional ambiguous or unambiguous displays, would then be presented at regular intervals throughout the block. Our prediction is that the responses to the test sequences would be influenced by the probability of objective streaming and bouncing events preceding the test: more frequent objective bounces prior to a test event should prime the participant to expect a bounce and hence a higher probability of a bounce response. Alternatively, more frequent streaming events prior to test should prime the participant's expectation the other way and result in a lower probability of a bounce response. We further predict that such prior events would modulate the influence of disambiguating factors such as introducing a vertical offset between target trajectories or rendering the targets visually distinguishable.

5. Conclusions

Our data are consistent with the idea that the perceptual system employs a form of inference based on regularities such as the basic laws of physics when interpreting a sensory event. In our experiments a collision is inferred as being more likely if preceding events are consistent with that interpretation.

References

Anstis, S. and Ramachandran, V. S. (1987). Visual inertia in apparent motion, *Vis. Res.* **27**, 755–764.

Bertenthal, B. I., Banton, T. and Bradbury, A. (1993). Directional bias in the perception of translating patterns, *Perception* **22**, 193–207.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spat. Vis.* **10**, 433–436.

Burns, N. R. and Zanker, J. M. (2000). Streaming and bouncing: observations on motion defined objects, *Clin. Exp. Ophthalmol.* **28**, 220–222.

Fujisaki, W., Shimojo, S., Kashion, M. and Nishida, S. (2004). Recalibration of audiovisual simultaneity, *Nat. Neurosci.* **7**, 773–778.

Grassi, M. and Casco, C. (2009). Audiovisual bounce-inducing effect: attention alone does not explain why the discs are bouncing, *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 235–243.

Grassi, M. and Casco, C. (2010). Audiovisual bounce-inducing effect: when sound congruence affects grouping in vision, *Atten. Percept. Psychophys.* **72**, 378–386.

Grassi, M. and Casco, C. (2012). Revealing the origin of the audiovisual bounce-inducing effect, *Seeing Perceiving* **25**, 223–233.

Grove, P. M. and Sakurai, K. (2009). Auditory induced bounce perception persists as the probability of a motion reversal is reduced, *Perception* **38**, 951–965.

Grove, P. M., Kawachi, Y. and Sakurai, K. (2012a). The stream/bounce effect occurs for luminance- and disparity-defined motion targets, *Perception* **41**, 379–388.

Grove, P. M., Ashton, J., Kawachi, Y. and Sakurai, K. (2012b). Auditory transients do not affect visual sensitivity in discriminating between objective streaming and bouncing events, *J. Vis.* **12**, 1–11.

Kawabe, Y. and Miura, K. (2006). Effects of orientation of moving objects on the perception of streaming/bouncing motion displays, *Percept. Psychophys.* **68**, 750–758.

Kawachi, Y. and Gyoba, J. (2013). Occluded motion alters event perception, *Atten. Percept. Psychophys.* **75**, 491–500.

Kawachi, Y., Grove, P. M. and Sakurai, K. (2014). A single auditory tone alters the perception of multiple visual events, *J. Vis.* **14**, 1–13.

Pelli, D. (1997). The Video Toolbox software for visual psychophysics: transforming numbers into movies, *Spat. Vis.* **10**, 437–442.

Sanabria, D., Correa, Á., Lupiáñez, J. and Spence, C. (2004). Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion, *Exp. Brain Res.* **157**, 537–541.

Sanabria, D., Lupiáñez, J. and Spence, C. (2007). Auditory motion affects visual motion perception in speeded discrimination task, *Exp. Brain Res.* **178**, 415–421.

Sekuler, A. B. and Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception* **28**, 415–432.

Sekuler, A. B., Sekuler, R. and Lau, R. (1997). Sound alters visual motion perception, *Nature* **385**, 308.

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions, *Curr. Opin. Neurobiol.* **11**, 505–509.

Watanabe, K. and Shimojo, S. (1998). Attentional modulation in perception of visual motion events, *Perception* **27**, 1041–1054.

Watanabe, K. and Shimojo, S. (2001a). Postcoincidence trajectory duration affects motion event perception, *Percept. Psychophys.* **63**, 16–28.

Watanabe, K. and Shimojo, S. (2001b). When sound affects vision: effects of auditory grouping on visual motion perception, *Psych. Sci.* **12**, 109–116.

Zhou, F., Wong, V. and Sekuler, R. (2007). Multisensory integration of spatio-temporal segmentation cues: one plus one does not always equal two, *Exp. Brain Res.* **180**, 641–654.