Perception of causality and synchrony dissociate in the audiovisual bounce-inducing effect (ABE)

Vroomen, Jean; Keetels, Mirjam

Perception of causality and synchrony dissociate in the audiovisual bounce-inducing effect (ABE)

Jean Vroomen*, Mirjam Keetels

Department of Cognitive Neuropsychology, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000, LE, Tilburg, the Netherlands

ABSTRACT

A sound can cause 2 visual streaming objects appear to bounce (the audiovisual bounce-inducing effect, ABE). Here we examined whether the stream/bounce percept affects perception of audiovisual synchrony. Participants saw 2 disks that either clearly streamed, clearly bounced, or were ambiguous, and heard a sound around the point of contact (POC). They reported, on each trial, whether they perceived the disks to 'stream' or 'bounce', and whether the sound was 'synchronous' or 'asynchronous' with the POC. Results showed that the optimal time of the sound to induce a bounce was before the POC (-59 msec), whereas audiovisual synchrony was maximal when the sound came after the POC ($+16$ msec). The range of temporal asynchronies perceived as 'synchronous', the temporal binding window (TBW), was wider when disks were perceived as bouncing than streaming, with no difference between ambiguous and non-ambiguous visual displays. These results demonstrate 1) that causality differs from synchrony, 2) that causality widens the TBW, and 3) that the ABE is perceptually real.

1. Introduction

Imagine two billiard balls rolling toward each other and then bouncing off, thereby producing a sound. In our mind, the visual bounce and the sound appear to be synchronous and causally related. This inference makes sense, because in many natural multisensory situations, synchrony and causation go hand-in-hand. However, sometimes there is a theoretically interesting exception to this rule, a case where synchrony and causality actually dissociate.

Many theorists have argued that temporal synchrony is of utmost importance for multisensory perception to occur. In order to perceive external events, the brain must solve the problem of how to integrate information from different sensory modalities whose signals can differ in time and space. For multisensory integration to occur, stimuli must fall in a specific range of temporal onsets, often termed the temporal binding window (TBW) (Stein & Meredith, 1993). Stimuli within the TBW (for audiovisual stimuli usually within the sub-second range of -100 msec (sound early) to $+200$ msec (sound late)) are generally perceived as being simultaneous and are then fused in an optimally weighted fashion to yield an integrated percept (Ernst & Banks, 2002). Further research has demonstrated that the TBW is quite plastic in nature: it decreases in size during development from infancy through adulthood (Lewkowicz, 1996), it can be shifted via exposure to leading or lagging stimulus combinations (Fujisaki, Shimojo, Kashino, &

Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004), it can be sharpened via training (Stevenson, Wilson, Powers, & Wallace, 2013), and its size is particularly large for audiovisual speech (for review, see Vroomen & Keetels, 2010).

It has also been argued that whenever stimuli fall within the TBW, they are not only perceived as being synchronous, but also as 'causally' related (Michotte, 1963; Wagemans, 2018). The critical idea is that temporal coincidence of multiple changes is unlikely to be a matter of chance, and a simultaneous occurrence of events is a non-accidental property that constitutes a basis for grouping them perceptually. In this view, then, events within the TBW are perceived as simultaneous, causally related, and are fused into a coherent optimally-weighted percept (e.g., Kording et al., 2007).

This monolithic view of the TBW, though, may need refinement because causation and synchrony may not always go hand-in-hand. In fact, for events to be causally related there is a logical necessity that cause-comes-before-consequence, and this serial order might, in principle, conflict with perception of synchrony as such. In fact, for perception of audiovisual synchrony, there is a general preference that a sound occurs *after* a visual event rather than before, possibly because audition has faster neural processing times than vision (Dixon & Spitz, 1980; Vroomen & Keetels, 2010). This sound-late bias for perceiving audiovisual synchrony implies that causation and synchrony might dissociate in cases where a sound causes a visual event.

A well-known example of the latter is the effect that a sound has on the stream-bounce illusion (i.e., the audiovisual bounce-inducing effect, ABE). In this illusion, two identical objects move steadily toward one another, coincide, and then move apart. This display is consistent with two different interpretations: either, after coincidence, the two objects could have continued in their original directions (streaming); or they could have collided and then bounced, reversing directions. In this ambiguous situation, a strong bias is found toward the perception of streaming. Sekuler, Sekuler, and Lau (1997) showed that the incidence of a bouncing percept could be increased by presenting a sound before (−150 msec), or at (0 msec) the point of contact (POC) of the two disks, whereas a sound after the POC (+150 msec) had significantly less effect, although it still enhanced perception of bouncing. The ABE has been taken as an example that the sound causes the ambiguous motion display to bounce, although a large number of studies have reported that other factors can induce a bounce as well, such as a momentary pause of the two disks, a salient flash, or a tactile vibration around the POC (Watanabe, 2001).

The ABE allowed us to examine a hitherto rather unexplored aspect of this illusion, namely whether a sound that makes the visual display appear to bounce is actually perceived as being synchronous (or asynchronous) with that bounce (or stream). One possibility already alluded to, is that the optimal timing for inducing a bounce occurs *before* the POC, whereas audiovisual synchrony is expected to be optimal when the sound occurs *after* the POC. The optimal timing of the sound for causality and synchrony might thus be different, so that a sound before the POC induces a bounce, but is nevertheless perceived as being asynchronous with that bounce. Results along this line have already been reported before by (Watanabe, 2001) who found that sounds at around −100 msec before the POC induced a bounce, whereas (4 trained) observers were, in separate sessions, nevertheless able to notice this audiovisual asynchrony. Here, we further explored this intriguing finding with an arguably more sensitive task, i.e., a simultaneity-judgement (SJ) task, instead of temporal order judgements. In an SJ-task, participants judge whether two events are simultaneous or not, instead of which came first/s. Previous research strongly suggests that the SJ task should be preferred over the temporal order judgement task when the primary interest is in perceived audio-visual synchrony (van Eijk, Kohlrausch, Juola, & van de Par, 2008). We also used a larger and more naïve sample of participants, assessed simultaneity and causality within the same trial rather than in separate sessions, and used displays that allowed us to compare sensitivity to ambiguous versus non-ambiguous (clear) bouncing and clear streaming displays.

A theoretically more intricate issue is that apparent causality might also actually change perception of synchrony. A well-known example from the motor domain, known as 'intentional binding', is that the perceived time of an intentional action (like a voluntary tap of the finger) and their sensory consequences (a sound) are attracted together in conscious awareness, so that subjects perceive their voluntary movements as occurring later and their sensory consequences as occurring earlier than they actually did (Haggard, Clark, & Kalogeras, 2002). A similar argument has been made for the audiovisual case by (Kohlrausch, van Eijk, Juola, Brandt, & van de Par, 2013). These authors used a visual display of a ball that apparently fell toward and bounced off a visible (or invisible) bar together with an impact sound, while subjects made synchrony judgements about the bounce and sound. The display either showed the full motion of the ball, the motion toward the bar with a sudden stop, or only the lift off. The visibility of the bar (i.e., whether the timing of the visual reversal or stop could be predicted or not) had no effect on synchrony judgements, but the lift off display was more tolerant for sound-early timings than the full animation and sudden stop (which were similar). The authors argued that the critical difference in the lift off display was a 'reversal of causality' so that the sound appeared to cause the lift off rather than that the visual bounce (or sudden stop) caused the sound. Apparent causation, rather than predictive information, was thus argued to modulate

perception of audiovisual synchrony. Importantly, though, the timing of the lift off could not be predicted in this situation, and the reversal in causality was only used as a post hoc explanation rather than being formally assessed.

Previous studies on the effects of audiovisual causality on synchrony judgements are therefore confounded because they do not have a measure of causality. Most relevant for the current case is that perception of audiovisual synchrony and causality have never been measured on a trial-by-trial basis with identical displays that measure causality and synchrony at the same time. In order to critically examine this, we used the ABE and asked participants, on each trial, to report whether the disks were streaming or bouncing (thus providing a measure of causation for ambiguous displays), and whether the sound was synchronous or asynchronous with the POC (thus providing a measure of audiovisual synchrony). This allowed us to separately analyze synchrony judgements of physically identical trials that were perceived as either streaming or bouncing. Most critically, if causality indeed widens the TBW, one would expect ambiguous motion displays perceived as bouncing to be more tolerant for sound delays (a larger TBW) than identical displays perceived as streaming.

Our study also allowed us to compare synchrony judgements in ambiguous motion displays perceived as streaming or bouncing with *non*-ambiguous streaming or non-ambiguous bouncing displays. These non-ambiguous displays were made by using two differently colored disks (black and white), instead of two identical ones, which streamed through or bounced of each other. This comparison between ambiguous and non-ambiguous displays is important because it is informative about the perceptual nature of the ABE itself. There has been much debate regarding the perceptual nature of cross modal illusions, including the ABE. For example, (Bertelson & de Gelder, 2004) have argued that one needs to take every precaution to try and rule out the possible confounding influence of response biases. For the ABE, one might argue that the effect of the sound simply reflects some kind of cognitive bias, rather than a genuine cross modal perceptual effect. While some attempt has been made to address this cognitive bias issue (Watanabe & Shimojo, 2001), this research still used subjective reports about streaming or bouncing. In order to overcome this potential methodological short-coming, we used a more indirect measure of perception, relying instead on a modulation of synchrony judgements. Most critically, if the ABE is perceptually 'real', one expects synchrony judgements of ambiguous motion displays (perceived as streaming or bouncing) to be like their non-ambiguous streaming or non-ambiguous bouncing counterparts.

2. Materials and methods

2.1. Subjects

A total of 27 students (19 females, age-range 18–30 years) from Tilburg University took part in the study and received course credits for their participation. Five of them were later discarded because a Gaussian function could not be fit on one of their response distributions, either because they gave too few stream (or bounce) response in ambiguous displays (4), or because they did not have an ABE (1). Participants reported normal hearing and normal or corrected-to-normal seeing. All participants were tested individually and were unaware of the purpose of the experiment. Written informed consent was obtained from each participant (in accordance with the Declaration of Helsinki). The Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University approved all experimental procedures (EC-2016.48a2).

2.2. Apparatus and stimuli

The experiment took place in a dimly lit and sound-attenuated chamber. Stimulus presentation was scripted using *E*-Prime 3.0
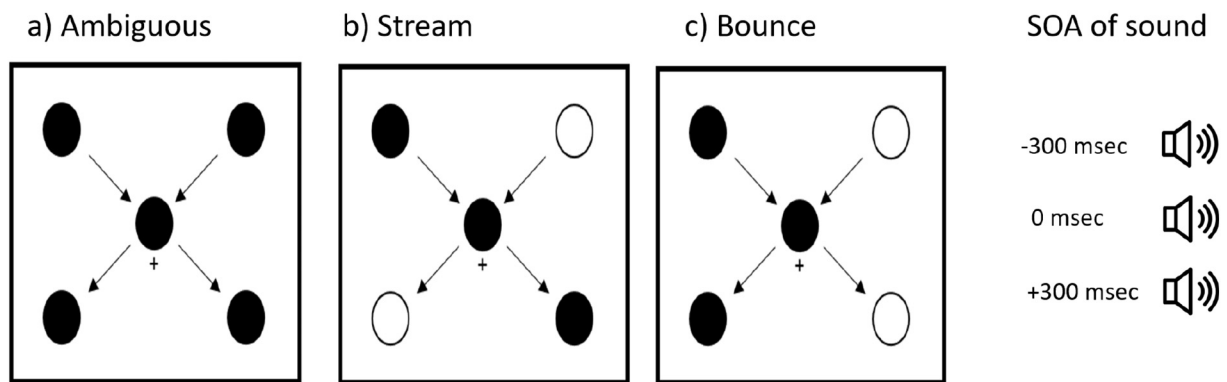
Visual motion displays

Fig. 1. Spatiotemporal dynamics of the three visual motion displays. Two disks moved toward each other, collided, and then moved apart. The color of the disk made the movement of the disks ambiguous, clearly streaming, or clearly bouncing. Sounds were presented around to point of contact (POC) at various sound onset asynchronies (SOA). The task of the participant was to judge, on each trial, whether 1) the disks were streaming or bouncing, and 2) whether the sound was synchronous or asynchronous at the POC.

software (Psychology Software Tools, Pittsburgh, PA). Visual stimuli were presented on a 24.5-in. LCD screen (BenQ Zowie XL2540, resolution 1920 × 1080, refresh rate 240 Hz). Auditory stimuli were delivered by two loudspeakers (Edifier R1280T) positioned at the left and right sides of the screen so that a single sound presented by the two speakers simultaneously would be perceived as coming from the center of the screen. Participants sat in front of the screen at a viewing distance of approximately 57 cm.

Visual stimuli consisted of two disks (each 0.8° in diameter) following a rectilinear trajectory in an oblique direction (see Fig. 1). On the first frame, the two disks were presented at the left and right of the screen (at 10.7° from the center) just above the center of the screen (at 1.4°). The two disks moved for 1.25 s at 17.6°/sec in an oblique direction with uniform rectilinear motion to a position at 10.7° to the left or right on the other side of the screen, 4.1° below the center. At the POC, the two disks completely overlapped at 1.4° below the center of the screen. The disks were presented on a gray background. A white fixation cross was presented at 2.4° below the center of the screen for the duration of a trial.

There were three visual motion displays: Ambiguous, Non-ambiguous streaming, Non-ambiguous bouncing. In the ambiguous displays, both disks were either white or black and did not change color during their motion paths, leading to perceptual ambiguity. In the clear streaming and bouncing displays, one of the disks was black and the other one was white (in half of the trials the left disk was white). In the bouncing display, the disk color changed after the POC, consistent with disks that bounced off each other. In the streaming display, the disk color did not change after the POC, consistent with disks that followed their initial trajectory and moved through each other.

The auditory stimulus was a brief white noise (62 dB, 9 ms), presented at one of nine different Stimulus Onset Asynchronies (SOA) relative to the POC (−300 msec, −150 msec, −75 msec, −37.5 msec, 0 msec, +37.5 msec, +75 msec, +150 msec, and +300 msec; negative values represent the sound before the POC). All SOAs were confirmed with an oscilloscope, with < 1 msec timing error. A visual-only condition was added that served as a baseline.

2.3. Procedure and design

Participants fixated on the fixation cross for the duration of a trial. A trial started with a blank screen for 500 msec, after which the disks appeared on the screen and started their trajectory. The sound, if present, was played at one of the nine possible SOA's. After the disks had

finished their trajectories, participants judged whether the disks were 'streaming' or 'bouncing' (SB-task), and then judged whether the sound, if present, was 'synchronous' or 'asynchronous' at the POC (SJ-task). Participants initiated the following trial by clicking a 'Next'-button.

The three visual displays (Ambiguous, Stream, Bounce) were presented randomly in seven blocks of 60 trials each (420 trials in total). Within a block, each of the nine SOAs and the visual-only condition were presented twice in random order. A training session of 60 trials preceded the experiment. Total testing time was about 60 min.

3. Results

Trials of the training session were excluded from further analysis.

3.1. SB-task

For the SB-task, the individual proportion of 'bounce'-responses was calculated for each type of motion display and SOA (see Fig. 2).

Across all SOAs, clear bouncing displays were almost always perceived as bouncing (p(Bounce) = 0.82), clear streaming displays as streaming (p(Bounce) = 0.05), and the silent ambiguous motion display was in between (p(Bounce) = 0.30), thus indicating that the visual displays were perceived as intended.

For ambiguous motion displays with sound, the presence of a sound increased, as expected, the proportion of bouncing responses depending on SOA (the ABE). For each individual, a Gaussian function was fitted on this distribution using the Matlab psignifit toolbox version 2.5.6 (Wichmann & Hill, 2001) to estimate the mean, amplitude, and standard deviation (SD). The group-averaged mean of the distribution, representing the optimal timing of the sound to induce a bounce, was at −59 msec, thus indicating that the optimal time of the sound to induce a bounce was well *before* the POC. The SD of the distribution, (representing sensitivity) was 222 msec, indicating that a sound induced a bounce at a rather wide range of SOA's (see Table 1). Separate *t*-test on the individual proportion of bounce responses per SOA showed that, relative to the silent baseline condition (p(Bounce) = 0.30), all sounds induced a bounce, except those at the extreme ends of the SOA (at −300 msec; +150 msec and +300 msec).

3.2. SJ-task

For the SJ-task, the individual proportions of 'synchronous' responses was calculated for each type of motion display (Ambiguous or
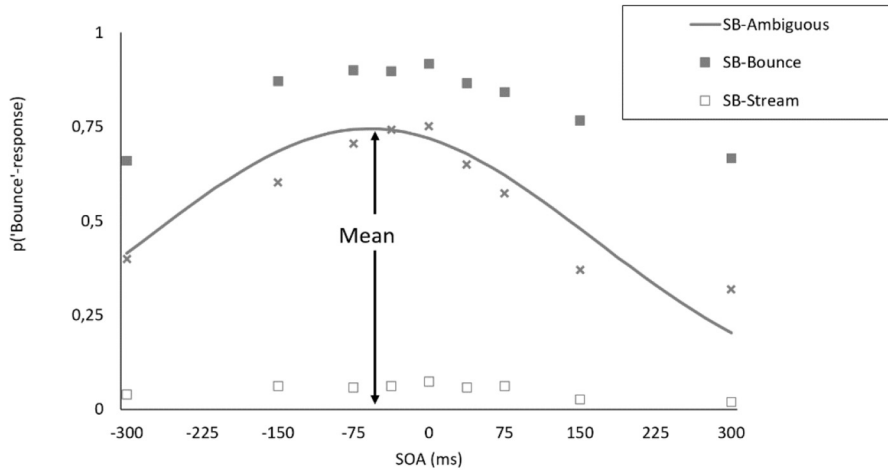
Stream/Bounce Task

Fig. 2. The average chance of responding 'bounce' for each of the three visual motion displays as a function of the SOA of the sound. Negative values represent sound-first stimulus pairs. The bell-shaped line represents the Gaussian fit on the ambiguous motion displays. The mean (at −59 msec) is the SOA at which the chance of responding 'bounce' is maximal. The solid horizontal line is the baseline of the silent visual condition.

Table 1

Averaged standard deviations (SD) and means of the individually fitted Gaussian functions on the Stream-Bounce (SB) task and the Simultaneity Judgement (SJ) task.

Task	Visual Motion Display	SD	(SEM)	Mean	(SEM)
SB	Ambiguous	222,4	(18.2)	−58,9	(10.8)
SJ	Ambiguous/stream percept	129,0	(11.9)	22,3	(5.8)
	Ambiguous/bounce percept	183,5	(19.9)	28,5	(10.0)
	Clear stream	130,1	(9.1)	3,3	(5.6)
	Clear bounce	165,3	(12.2)	32,0	(7.9)

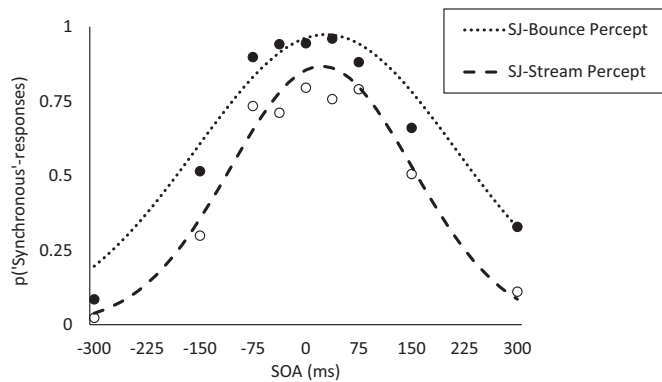## SJ-Task: ambiguous display perceived as Bounce or Stream

Fig. 3. The average chance of responding 'synchronous' for ambiguous motion displays perceived as either streaming or bouncing, as a function of the SOA of the sound. Negative values represent sound-first stimulus pairs. The dotted lines represent the Gaussian fit. Displays perceived as bouncing had a wider range of SOAs perceived as being synchronous with the sound (the temporal binding window (TBW)) than identical displays perceived as streaming.

Non-ambiguous), percept (Streaming or Bouncing) and SOA (see Figs. 3 and 4). For each participant and condition, Gaussian functions were then fitted across SOAs. The mean of the Gaussian represents the Point of Subjective Simultaneity (PSS), which is the SOA at which the sound is perceived to be maximally synchronous. The standard deviation of the Gaussian fit is a measure of sensitivity to audiovisual (a)synchrony and it represents the theoretically important TBW (see Table 1).

A 2 (Display type: Clear or Ambiguous) x 2 (Percept: Streaming or Bouncing) ANOVA on the TBWs showed that the TBW of displays perceived as bouncing were, on average, 45 msec wider than displays perceived as streaming (174 msec vs. 129 msec, respectively; $F_{(1,21)} = 22.79$, $p < .001$). There was no significant difference

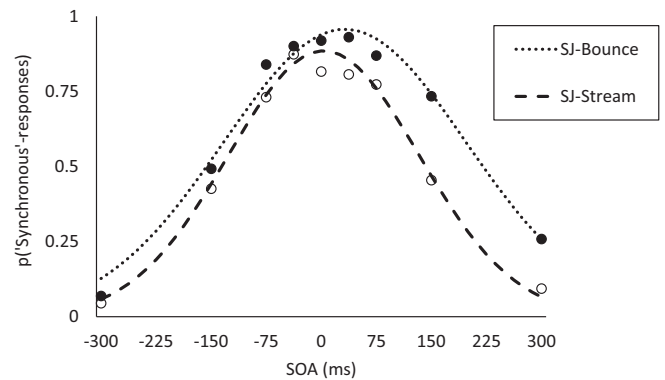## SJ-Task: non-ambiguous Bounce and Stream displays

Fig. 4. The average chance of responding 'synchronous' for non-ambiguous streaming and non-ambiguous bouncing displays. The dotted lines represent the Gaussian fit. Non-ambiguous bouncing displays had a wider range of SOAs (the temporal binding window (TBW)) than non-ambiguous streaming displays.

between ambiguous versus non-ambiguous displays, $F_{(1,21)} = 0.95$, $p = .34$, and no interaction between the two factors, $F_{(1.21)} = 1.39$, $p = .25$. Bouncing displays, ambiguous and non-ambiguous alike, thus had a wider TBW than streaming displays.

A similar 2 (Display type: Clear or Ambiguous) × 2 (Percept: Streaming or Bouncing) ANOVA was run on the individual PSSs., The PSS of displays perceived as bouncing was +17 msec later than displays perceived as streaming (+30 vs. +13 msec, respectively; $F_{(1,21)} = 8.79$, $p < .007$), with no main effect of display type, $F_{(1,21)} = 2.60$, $p = .12$, and no interaction, $F_{(1.21)} = 2.86$, $p = .10$. An overall ANOVA on the intercept (testing against 0) showed that the grand average PSS was well *after* the POC (at +21 msec) reflecting an overall sound-late bias for perception of audiovisual synchrony, $F_{(1,21)} = 14.72$, $p < .001$. Perception of audiovisual synchrony was thus maximal when the sound occurred *after* the POC, and displays perceived as bouncing (clear and ambiguous alike) had a later PSS than displays perceived as streaming (clear and ambiguous alike), thus again without difference between ambiguous and non-ambiguous motion displays.

3.3. Comparison of the SB- and SJ-task

The results of the SB and SJ task indicate that the optimal time of a sound to induce a bounce is earlier than the optimal timing to perceive audiovisual synchrony. To formally test this, we compared the means of

the Gaussian functions of the SB and SJ task in a paired *t*-test. It showed that the 75 msec difference was indeed significant (t(21) = 7.3, p < .001).

4. Discussion

Here we demonstrate, for the first time, that causally related audiovisual events have a wider temporal binding window (TBW) than physically identical events perceived as being causally unrelated. Using the audiovisual bounce effect (ABE), participants were asked on each trial to judge whether motion displays were streaming or bouncing, and to judge whether the sound was synchronous or asynchronous at the point of contact (POC). The results demonstrate, firstly, that sounds in *ambiguous* motion displays perceived as bouncing had a wider TBW (and a later PSS) than identical displays perceived as streaming. Furthermore, *non-ambiguous* motion displays were perceived in a similar way: clear bouncing displays had a wider TBW (and a later PSS) than clear streaming displays. This is important because it the first clear demonstration that perception of causality, when actually measured at single trial level with identical stimuli, indeed widens the temporal delays at which sounds are perceived as being synchronous. This result aligns with previous reports on this matter (Haggard et al., 2002; Kohlrausch et al., 2013).

Secondly, the optimal time of the sound to induce a bounce differs from the optimal time to induce audiovisual synchrony. The ABE was maximally effective when the sound was presented *before* the POC (Sekuler et al., 1997), whereas perception of audiovisual synchrony was maximal when the sound was presented *after* the POC (Vroomen & Keetels, 2010). A bounce-inducing sound could thus be perceived as being asynchronous with that bounce (see also Watanabe, 2001). An analogous finding has been reported with the McGurk illusion (McGurk & MacDonald, 1976) where multisensory integration and audiovisual synchrony can diverge. In one particular example of this illusion, participants may be presented with a face articulating 'ba' while hearing the sound 'da', resulting in a blended percept 'bda'. In a study using this illusion (Soto-Faraco & Alsius, 2007), it was found that when participants were presented with these stimuli at various SOAs and asked to report both their percept and the temporal order of the stimuli, these two types of judgements did not necessarily line up. At some SOAs, participants correctly noticed that the sound 'da' was presented before video 'ba', but nevertheless perceived a 'bda', so with phoneme order reversed. This suggests that speech and non-speech stimuli can be integrated in a multisensory percept despite noticeable delays between the auditory and visual input streams.

Thirdly, audiovisual synchrony judgements in ambiguous motion displays perceived as bouncing were like their clear bouncing counterpart (a wider TBW), while ambiguous motion displays perceived as streaming were like their clear streaming counterpart (a smaller TBW). This correspondence between clear and ambiguous motion displays underlines the perceptual reality of the ABE (see also Meyerhoff, 2018).

It remains for future research to determine what the unifying mechanism is that a sound not only induces a bounce, but also widens the TBW. For the ABE, it has been argued that a transient stimulus like a sudden sound (or tap/flash) distracts visual attention, and that this temporary lack of attention around the POC induces a bouncing percept (Watanabe, 2001). Could it be that this lack of attention also automatically widens the TBW? Previous research on the role of attention (or the lack thereof) on audiovisual simultaneity judgements provides a mixed picture. On the one hand, it appears that multisensory synchrony judgements of transient stimuli like flashes, taps, and beeps are quite immune to a lack of attention (Vroomen & Keetels, 2010), whereas the results for moving/bouncing stimuli, as in the ABE, are mixed. A study by (Donohue, Green, & Woldorff, 2015) asked participants to judge audiovisual synchrony of streaming/bouncing stimuli that appeared at spatially attended or unattended sides. They found that endogenous visuo-spatial attention only increased the proportion of synchronous

responses when sounds were actually synchronous at the POC, but attention did not change synchrony judgements in sound-late trials. The size of the TBW in the ABE thus seemed to be quite immune to a lack of spatial attention. (Note that these authors did not include sound-early trials, and did not analyze data separately for streaming versus bouncing percepts, as was critical here).

If not a temporary distraction of attention, what else could account for the ABE and the widening of the TBW? It is known that judging audiovisual simultaneity, as well as the ABE, mainly depend on the availability of clear temporal onsets in vision and audition (Meyerhoff & Suzuki, 2018; Vroomen & Keetels, 2010). At least introspectively, a bounce - whether induced via the ABE or by non-ambiguous bouncing display -has a clearer visual temporal onset than a stream, and this difference might contribute to a widening of the TBW. An event-related functional magnetic resonance imaging (fMRI) study has shown that a transient sound induced higher activation in multimodal areas (e.g., the prefrontal and posterior parietal cortex) and subcortical areas (thalamus) when participants had a bouncing rather than streaming percept (Bushara et al., 2003). The authors interpret this activation pattern as evidence for a reciprocal and competitive interaction between multimodal and predominantly unimodal processing networks. Possibly, these reciprocal interactions in multisensory areas also underlie the widening of the TBW.

5. Conclusion

The law of perceiving causality (cause-before-consequence) can conflict with the law of perceiving synchrony. This is apparent in the audiovisual bounce illusion where an early sound causes the two disks to a bounce, but is nevertheless perceived as being asynchronous with that bounce. This often goes unnoticed, though, because causally related events are more tolerant to temporal asynchronies than identical events perceived as causally unrelated.

Author contributions

JV and MK conceived and designed the study; MK collected and analyzed the data; JV interpreted the results and wrote the manuscript.

Declaration of competing interest

The authors declare no competing interests.

References

Bertelson, P., & de Gelder, B. (2004). The psychology of multisensory perception. In C. Spence, & J. Driver (Eds.). *Crossmodal Space and Crossmodal Attention* (pp. 141–177). Oxford University Press.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience, 6*(2), 190–195. https://doi.org/10.1038/nn993.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719–721.

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience, 9*(APR), https://doi.org/10.3389/fnint.2015.00032.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. https://doi.org/10.1038/415429a.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*, 773–778.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382–385.

Kohlrausch, A., van Eijk, R., Juola, J. F., Brandt, I., & van de Par, S. (2013). Apparent causality affects perceived simultaneity. *Attention, Perception, and Psychophysics, 75*(7), 1366–1373. https://doi.org/10.3758/s13414-013-0531-0.

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One, 2*(9), https://doi.org/10.1371/journal.pone.0000943 ARTN e943DOI.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology-Human Perception and Performance, 22*(5),

1094–1106.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

Meyerhoff, H. S. (2018). *Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents. September 2017*https://doi.org/10.1016/ j.cognition.2017.08.007.

Meyerhoff, H. S., & Suzuki, S. (2018). Beep, Be-, or -Ep: The impact of auditory transients on perceived bouncing/streaming. *Journal of Experimental Psychology: Human Perception and Performance, 44*(12), 1995–2004. https://doi.org/10.1037/ xhp0000585.

Michotte, A. (1963). The perception of causality. https://psycnet.apa.org/record/1964-05029-000.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308. https://doi.org/10.1038/385308a0.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport, 18*(4), 347–350. http://www.ncbi.nlm.nih.gov/ entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids= 17435600.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* The MIT Press.

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research,*

225(4), 479–489. https://doi.org/10.1007/s00221-012-3387-y.

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics, 70*(6), 955–968. https://doi.org/10.3758/pp.70.6. 955.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention Perception & Psychophysics, 72*(4), 871–884.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32–35.

Wagemans, J. (2018). Perceptual organization. *The Stevens' handbook of experimental psychology and cognitive neuroscience: Vol. 2. Sensation, perception & attention. Vol. 2. The Stevens' handbook of experimental psychology and cognitive neuroscience: Vol. 2. Sensation, perception attention* (pp. 1–70). Hoboken, NJ: John Wiley & Sons, Inc.

Watanabe, K. (2001). *Crossmodal interaction in humans.*

Watanabe, K., & Shimojo, S. (2001). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics, 63*(1), 16–28. https://doi.org/10.3758/ BF03200498.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 43. http://users.ox.ac.uk/ ~sruoxfor/psychofit/.