View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Birmingham Research Portal

University of Birmingham Research at Birmingham

Causality shifts the perceived temporal order of audio-visual events

Fornaciai, Michele; Di Luca, Max

DOI:
10.1037/xhp0000754

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Fornaciai, M & Di Luca, M 2020, 'Causality shifts the perceived temporal order of audio-visual events', *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000754

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
©American Psychological Association, 2020. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: 10.1037/xhp0000754

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 14. Jun. 2020

Causality shifts the perceived temporal order of audio-visual events

Michele Fornaciai [1*], Massimiliano Di Luca [2]

[1] Cognitive Neuroscience Department, International School for Advanced Studies (SISSA), 34136 Trieste (TS), Italy.

[2] Centre for Computational Neuroscience and Cognitive Robotics, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

* Corresponding Author.
Michele Fornaciai
International School for Advanced Studies (SISSA),
Via Bonomea, 265
34136 Trieste (TS), Italy
e-mail: michele.fornaciai@gmail.com

Word count: 8048 words

ABSTRACT

Causality poses clear constraints to the timing of sensory signals produced by events, as sound travels slower than light, causing auditory stimulation to lag visual stimulation. Previous studies show that implied causality between unrelated events can change the tolerance of simultaneity judgements for audio-visual asynchronies. Here, we tested whether apparent causality between audio-visual events may also affect their perceived temporal order. To this aim, we used a disambiguated stream-bounce display, with stimuli either bouncing or streaming upon each other. These two possibilities were accompanied by a sound played around the time of contact between the objects, which could be perceived as causally related to the visual event according to the condition. Participants reported whether the visual contact occurred before or after the sound. Our results show that when the audio-visual stimuli are consistent with a causal interpretation (i.e., the bounce caused the sound), their perceived temporal order is systematically biased. Namely, a stimulus dynamics consistent with a causal relation induces a perceptual delay in the audio component, even if the sound was actually presented first. We thus conclude that causality can systematically bias the perceived temporal order of events, possibly due to expectations based on the dynamics of events in the real world.

Keywords: temporal order; causality; time perception; multisensory perception.

SIGNIFICANCE STATEMENT

- Causality represents a powerful cue to interpret perception, as the effect must always follow its cause. Here we tested whether such a cue provided by causality can also constrain the perception of arbitrary sensory events only *apparently* linked by a cause-effect relation.
- Our results suggest that implied causality between two otherwise unrelated events can distort their apparent timing, pushing the effect (an auditory signal) to be perceived after its apparent cause (a bounce between two objects), even if it was actually presented first.

1. INTRODUCTION

Time is one of the fundamental dimensions that the brain must process to achieve a correct representation of the external world and guide our behavior. One of the most important aspects of time is the temporal order of the events, which allows us to understand the structure of the environment and the relations between different events. Interestingly, temporal order perception has been shown to not always provide a precise and exact representation of the physical temporal order, but instead it appears to be an approximation that can deviate from veridicality. For instance, biases in perceived temporal order might be due to differences in processing latencies depending on the modality of the stimulus (Allison et al., 1983), or on its physical properties, like intensity (Matteson, 1971). Our tolerance to small mismatches in the timing of auditory and visual information is also biased, as small audio-lagging asynchronies very often pass unnoticed while audio-leading asynchronies are much more detectable (e.g., Dixon & Spitz, 1980; Munhall et al., 1996; Grant et al., 2004). This is likely due to the fact that in the natural environment sounds are more likely to come after visual information, *especially* when sound and light are linked by a cause-effect relation. For this reason, we posit that the perception of a causal relation between a visual and an auditory event may pose an additional constraint to their perceived temporal order.

Indeed, expectations based on the statistics of the external world have been shown to represent a strong prior for perception, increasing the efficiency of perceptual processing (i.e., speeding up the detection time, or increasing accuracy; Stein & Peelen, 2015; Pinto et al., 2015) and even altering how we perceive a stimulus (i.e., its appearance; Sterzer et al., 2008; Sotiropoulus et al., 2011). Such influences are especially powerful when stimuli are weak or uncertain, and have been linked to active and predictive processes interpreting the visual input in light of previous knowledge (e.g., see De Lange et al., 2018 for a review). These mechanisms of perceptual processing are well captured by Bayesian models of perception (e.g., Knill & Richards, 1996; Mamassian et al., 2002), which incorporate prior knowledge as a fundamental building block of perception. As causality represents a pervasive feature of the external environment, it should thus provide a prior to interpret the dynamics of audio-visual events.

Several studies investigated the influence of contextual information and causality on simultaneity and temporal order perception. For instance, Van Eijk et al. (2010) exploited experimental stimuli resembling a Newtons's cradle toy, displaying the full dynamics of the moving balls or covering one side of the cradle, in order to specifically provide only predictive or postdictive information. Results show that covering only the final part of the Newton's cradle animation did not change the perceived synchrony of the audio and visual events compared to showing the full movie. However, when the initial part of the animation was covered, perceived simultaneity substantially differed from the previous two cases. Specifically, perceptual performance measured with a simultaneity judgment (SJ) task showed a much steeper auditory-first synchrony boundary (i.e., characterizing the transition between perceived asynchrony to perceived synchrony), reflecting an increased sensitivity to audio-leading stimuli. These manipulations (i.e., covering the initial or final part of the cradle) effectively provided opposite causal interpretations, with the impact of the ball causing the sound in one case, or the sound being perceived as causing the launch of the rightmost ball, in the other case. These findings thus raise the possibility that the availability of visual information suggesting different causal interpretations may modulate the sensitivity to audio-visual asynchrony, in line with the idea that expectations based on the dynamics of events in the real world can systematically modulate perception. Moreover, results from van Eijk et al. (2010) also show that manipulating the availability of visual information and the related causal interpretation of the stimuli, has an impact also on PSS estimates obtained with a temporal order judgment (TOJ) task. In this case, PSSs are shifted toward the steeper decision boundary (i.e., for instance, the average TOJ PSS was shifted towards audio-leading asynchronies when the initial part of the cradle toy was covered).

In a subsequent study, to rule out the possibility that the observed results might be due to predictive information, Kohlrausch et al. (2013) showed stimuli falling down and bouncing on a bar, which could have been visible or invisible, accompanied by a sound. Critically, when the bar was invisible, no predictive information was available, while causality was still evident (the bounce caused the sound). What they found indicates that the changes in subjective simultaneity were explained completely by changes in the implied causal relation between audio and visual signals, with no influence of visual predictive information.

Here, we investigate whether an apparent cause-effect relation between an auditory and a visual event could shift their perceived *temporal order*. According to the idea that expectations can bias how we perceive a stimulus, our prediction is that a prior as strong as causality (i.e., due to its pervasiveness in the external environment) may not only alter the sensitivity to audio-visual simultaneity, but it could also directly alter the perceived relative timing of two events, shifting their apparent onset according to the expected dynamics of causally-related events. In this context, while previous results provide some hints suggesting a relation between apparent causality and perceived timing, whether causality can effectively alter the perceived temporal order of two events is unclear. For instance, results from van Eijk et al. (2008) show a relative perceptual delay of the auditory signal for bouncing-ball stimuli compared to flash-click pairs. However, such an effect was very small and not statistically significant, and confounded by the presence of predictive information in the bouncing ball condition. Similarly, results from van Eijk et al.'s (2010) study show a difference between audio-visual pairs entailing opposite causal interpretations – suggesting a role of causality in perceived timing – although again with the confounding factor of predictive information. Finally, as mentioned above, Kohlrausch et al. (2013) found more robust evidence for a role of causality on perceived timing. However, results from this study are limited to perceived simultaneity due to the specific task used (three-alternative simultaneity judgment task; SJ3). Indeed, it is important to note that different tasks requiring judging the temporal order of the stimuli (TOJ) or their synchrony/asynchrony (SJ and SJ3), yield different estimates of the point of subjective simultaneity, and may involve different underlying perceptual processes (e.g., van Eijk et al., 2008).

In order to more precisely assess the relation between causality and the perceived temporal order of audio-visual events, we thus employ a TOJ task. To assess the role of causal information in temporal order estimates, we use an unambiguous stream-bounce display by presenting two objects of different color moving towards each other (two circles in Exp. 1; one circle and one square in Exp. 2). Once the two objects reached the midline of the screen, two possible events could happen: the two objects streamed through each other without interacting, or the two objects bounced on each other, each changing its motion direction (or *appeared* to stream or bounce, in Exp. 2; see *Materials and Methods*). Around the time at which the two objects reached each other, a brief sound was presented. Participants were asked to indicate which one of the events (visual or auditory) happened first. Results across two experiments show that apparent causality between a visual and an auditory event (i.e., in the case of bouncing stimuli) systematically biases the perceived temporal order of the two events: the auditory event tended to be perceived after the visual one, even if the sound was actually presented first. This finding suggests that causality represents a powerful prior for the timing of events and it is capable of distorting their perceived temporal order, keeping perception consistent with expectations about the dynamics of events in the real world.

2. EXPERIMENT 1

2.1. Materials and methods

Subjects

A total of 12 participants (8 females; age ranging from 18 to 25 years old; including the author M.F.) took part in Experiments 1, after giving their written informed consent. They were compensated for their time with course credits or with a payment of 6 GBP/hour. All participants had normal or corrected-to-normal visual acuity and reported to have good hearing. With the exception of the author MF, all participants were naïve to the purpose of the experiment. Experiments were conducted according to the protocol approved by the STEM ethics' committee of the University of Birmingham, and are in line with the declaration of Helsinki. Note that to determine the sample size included in the experiment, we performed a power analysis taking into account previous research on the same topic (Kohlrausch et al., 2013). Namely, we took the PSS and SD measures from two of the conditions compared in Kohlrausch et al.'s study (i.e., the full motion and lift off stimuli in the visible bar condition), and estimated an effect size (Cohen's *d*) of about 1.4. However, to be more conservative, we took an estimated effect (*d*) size equal to 1. By considering such an approximate effect size, a power of 0.95, and a one-tailed distribution (i.e., according to our specific prediction), we estimated a required sample size equal to 13 participants. Even if Kohlrausch et al. (2013) used very different stimuli and task, this nevertheless provided a rough estimate of the number of participants required to obtain robust results in a similar context.

Apparatus and stimuli

Participants were tested in a quiet and dark room. Observers sat in front of a monitor screen at a distance of about 60 cm, with head stabilized by means of a chinrest. Stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) for Matlab (version r2013b; The Mathworks, Inc.) and they were presented on a CRT screen (NEC MultiSync FP2141sb; 40 x 30 cm) with a resolution of 800 x 600 and a refresh rate of 120 Hz. Visual stimuli were a pair of circles, with radius = 0.4 deg. One of the objects was presented in white, while the other in black, with such colors randomly assigned on each trial. Stimuli were displayed in two different colors to allow a better discrimination between the "stream" and "bounce" condition at the time of the event, making our stream-bounce display completely unambiguous. A brief sound was presented (frequency = 1,000 Hz, 62 dB, 9 ms) at one of eleven possible values of stimulus onset asynchrony (SOA) with respect to the moment at which the two objects contacted each other (-350, -150, -100, -66, -33, 0, 33, 66, 100, 150, 350 ms, where negative values represent audio-leading pairs). Sounds were played by a speaker (Fostex PM0.4n) juxtaposed to the left side of the screen. Synchrony between audio and visual signals was thoroughly calibrated before the experiment by means of an oscilloscope.

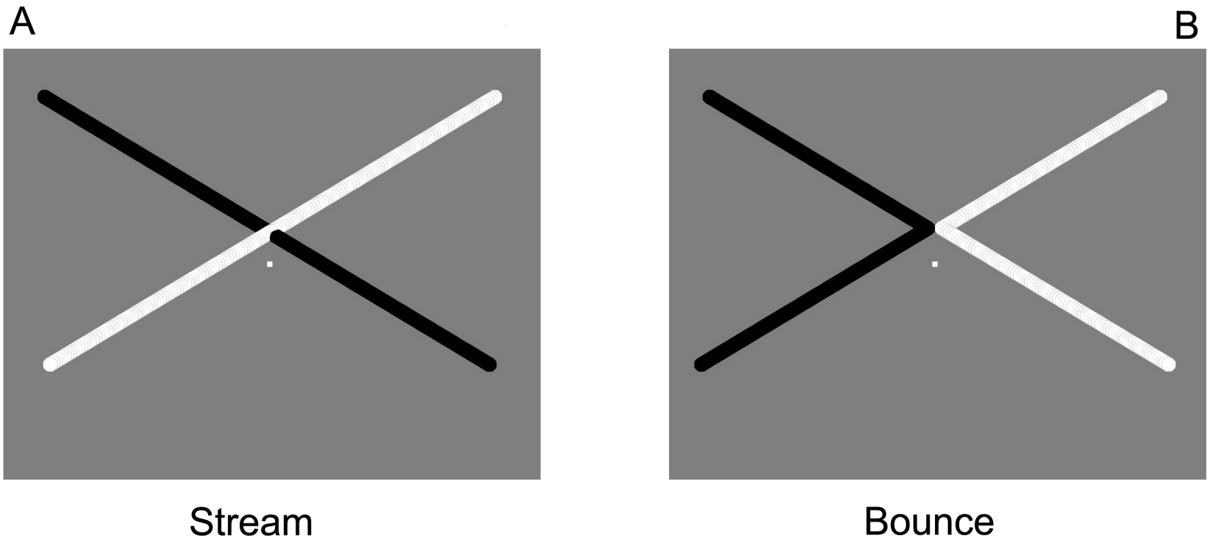


FIGURE 1 – Spatio-temporal dynamics of stimuli in Experiment 1. *Spatiotemporal dynamics of the stream (A) and bounce (B) conditions of Experiment 1. Note that stimuli are not depicted in scale.*

Procedure

Each trial started with subjects fixating a white dot at the center of the screen. After 500 ms, the two objects appeared on the screen (initial positions: 10 deg on the left and on the right of the central vertical midline, with a vertical eccentricity of 5 deg above the horizontal midline), and started to move towards each other, with a variable speed (8-10 deg/sec). After a variable amount of time, according to their speed, the two objects met upon the vertical midline of the screen, just slightly above the central fixation point (vertical eccentricity = 1 deg from the fixation point). At this point, the two objects could either stream through each other, continuing on their movement unperturbed, or bounce on each other, changing movement direction. This procedure allowed us to modulate the stimuli in a completely unambiguous fashion (i.e., the two events were clearly discriminable), but without providing visual anticipatory information (i.e., participants could not predict which event was presented in a given trial before the event itself). The two conditions were randomly intermixed within each block of trials. Participants were instructed to consider the moment at which the borders of the two objects touched as the visual event. Around the time of the visual event, a brief sound was presented, with one of the possible SOAs, randomized across trials. After the visual event, the two objects continued their motion until they disappeared at a position symmetrical to the starting point. At the end of each trial, participants were instructed to indicate which event (auditory or visual) happened first in the sequence, pressing the appropriate key ("A" for auditory first, "V" for visual first) on a keyboard. Pressing the key automatically triggered the start of the next trial (inter-trial interval = 650 ms). Most of the participants completed six blocks of 44 trials, with the exception of two subjects who completed five blocks of 44 trials due to the time constrains of the experimental session. A depiction of the stimulus trajectories is shown in Figure 1.

Participants' responses (in terms of proportion of visual-first judgments) as a function of physical audio-visual asynchrony were fitted with a cumulative Gaussian function, according to the Maximum likelihood method (Watson, 1979). The point of subjective simultaneity (PSS) was calculated as the median of the best-fitting cumulative Gaussian function to all the data of each participant and condition. As a measure of sensitivity to audio-visual asynchrony (or precision) we used the Just Noticeable Difference (JND), representing the standard deviation of the underlying Gaussian function. Additionally, in the psychometric fitting procedure, a finger error rate correction (5%) was applied in order to control for lapses of attention of response errors independent from the stimuli (Wichmann & Hill, 2001).

2.2. Results

Panels A and B of Figure 2 summarize the results of Experiment 1, showing the average participants' performance in terms of the point of subjective simultaneity (PSS) and precision (JND). As shown in panel 2A, in the bounce condition the auditory stimulus needed to be presented significantly earlier (PSS = -87 ± 20 ms) to match the perceived timing of the visual event, compared to the stream condition where the PSS was reached with asynchronies much closer to zero (PSS = -37 ± 26 ms; Wilcoxon Signed Rank test, bounce versus stream condition; Z = -2.27, p = 0.02, *d* = 0.58; note that we used a non-parametric test because the distribution of PSSs failed a normality test). The average difference between individual PSSs in the stream and bounce condition was -50 ms ± 25 ms. On the other hand, we did not observe any significant modulation of JND (Fig. 2B; paired t-test, t(11) = 1.05, two-tailed p = 0.32; JND = 68 ± 14 ms and 57 ± 7 ms, respectively for the stream and bounce condition; average difference = 11 ms ± 12 ms), suggesting that the different dynamics of the stimuli did not affect the precision of TOJ.

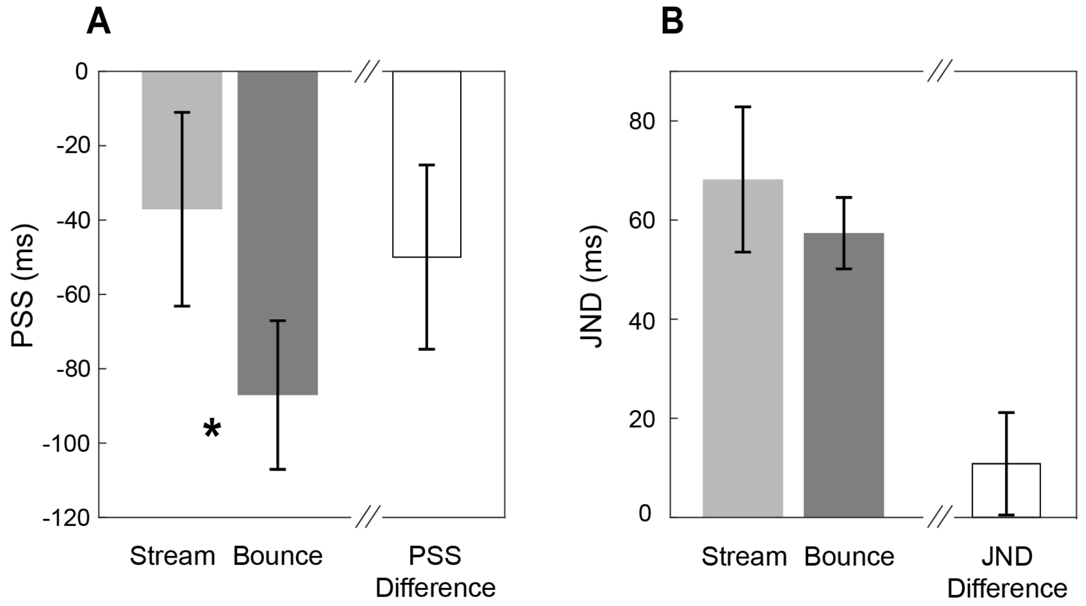


FIGURE 2 – Average PSSs and JNDs in Experiment 1. *Results of Experiment 1, in terms of average PSS (A) and JND (B). Negative values represent audio-leading stimulus pairs. PSS in the streaming and bouncing conditions resulted to be significantly different, with perceived simultaneity reached with more pronounced audio-leading asynchronies in the bouncing condition. JNDs, on the other hand, did not differ. The white bar in the rightmost part of both panels indicate the average individual difference between PSS or JND in the two conditions. Error bars are S.E.M. * $p < 0.05$.*

3. EXPERIMENT 2

Results from Experiment 1 suggest that apparent causality shifts the perceived temporal order of two events by either: (a) increasing the perceptual latency of the auditory signal or by decreasing the perceptual latency of the visual one in the bounce condition, or (b) by decreasing latency for audio or increasing latency for visual in the stream condition. The possibility of attributing the change in PSS to the visual stimulus (whether decreasing it in the bounce condition or increasing it in the stream condition), however, is more likely because of the specific visual sequence presented in Experiment 1, whereby the image of the circles in the streaming and bouncing conditions was different right after the contact. Namely, while in the stream condition the two circles continued on their own trajectory, thus superimposing upon each other before moving away, in the bounce condition the two stimuli changed trajectory immediately after the visual event. Even if the participants were instructed to consider the timing of the visual event to be the moment at which the two objects come into contact with each other, there is still the possibility that the visual difference after such a moment might have influenced the temporal order judgments. In this scenario, it is possible that the superimposition of the two objects in the stream condition could have been implicitly considered as the visual event, increasing the apparent perceptual latency of vision and leading to less negative PSSs in the stream condition.

To avoid such a confounding factor, in Experiment 2 we eliminated any possible difference between the two conditions. To do so, we presented intermittent stimuli visible in the same location along their trajectories for two frames and turned invisible for six frames (Figure 3). With this type of stimuli, the visual event (the joining of the borders) was correctly displayed, and then followed by a six-frames blank period. Presenting intermittent stimuli allowed us to display sequences where the position and appearance of the stimuli was identical, except that in half of the trials we swapped the two objects immediately after the visual event (during the six-frames blank). The two objects also differed in shape and color to enhance their discriminability

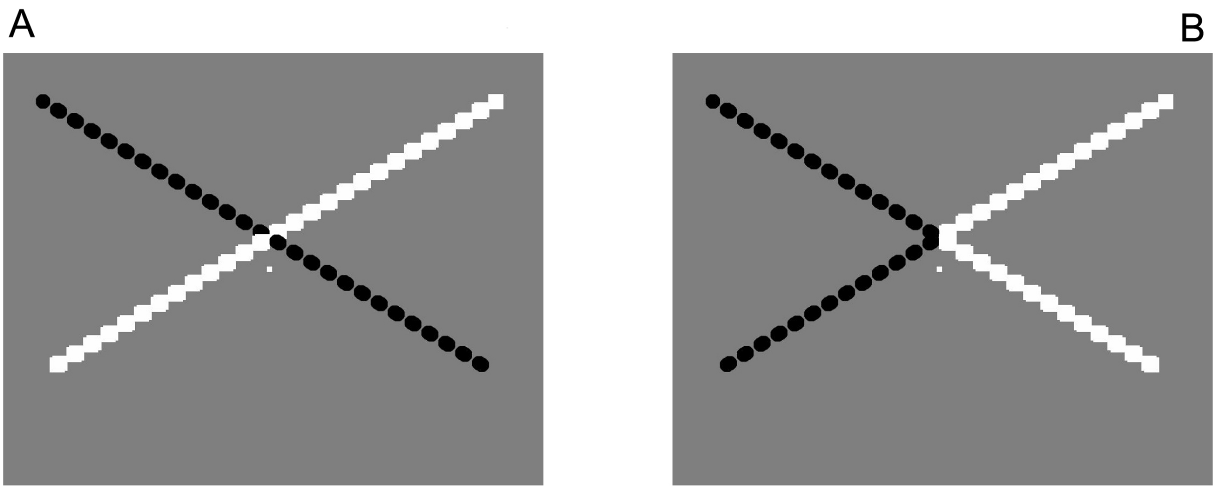
3.1. Materials and methods

Subjects

A total of 13 participants (11 females; age ranging from 18 to 25 years old; including the author M.F.) took part in Experiment 2, after providing written informed consent, and were compensated for their time with course credits or with a payment of 6 GBP/hour. All participants had normal or corrected-to-normal visual acuity and reported to have good hearing. With the exception of the author MF, all participants were naïve to the purpose of the experiment. Experiments were conducted according to the protocol approved by the STEM ethics' committee of the University of Birmingham, and are in line with the declaration of Helsinki. The sample size of this experiment was determined as in Experiment 1. In addition to the main conditions of Experiment 2, 11 more participants (10 naïve subjects and the author MF; age ranging from 19 to 34 years) were recruited to participate in the control causality perception task.

Apparatus and stimuli

The apparatus used in Experiment 2 was identical to Experiment 1. Visual stimuli were a circle (radius = 0.4 deg) and a square (0.4 x 0.4 deg), with one object presented in white and the other in black, according to a random selection performed in each trial. Again, stimuli were displayed in two different colors and two different shapes to allow a better discrimination between the “stream” and “bounce” events, making our stream-bounce display completely unambiguous at the time of the event (but not before the event itself). As in Exp. 1, a brief sound was presented around the time of contact of the two objects. The auditory stimulus and the level of SOAs compared to the visual event were identical to Exp. 1.



Stream

Bounce

FIGURE 3 – Spatio-temporal dynamics of stimuli in Experiment 2*. In Experiment 2, we devised intermittent stimuli and presented only streaming trajectories, in order to make the dynamics of the stream (A) and bounce (B) events identical. Such procedure indeed allowed us to blank the portion of the trajectory just after the visual event, so that in the bouncing condition we swapped the identities of the objects in order to simulate a bounce. Stimuli are not depicted in scale.*

Procedure

The procedure was largely similar to Experiment 1. Each trial started with subjects fixating on a central fixation point. After 500 ms from the trial start, the two objects appeared on the screen (initial positions: 10 deg on the left and on the right of the central vertical midline, with a vertical eccentricity of 5 deg above the horizontal midline), and started to move towards each other at 10 deg/sec. After about 1.2 s, the two objects met upon the vertical midline of the screen slightly above the central fixation point (eccentricity = 1 deg). Differently from Experiment 1, in Experiment 2 the stimuli were displayed intermittently only for two consecutive frames (~17 ms) with blank intervals of six frames (~50 ms), resulting in an intermittent presentation with stimuli moving in steps of approximately 0.5 deg. This manipulation, coupled with the appropriate object velocity (10 deg/s), allowed to keep the movement trajectories identical, with the bouncing event created by switching the color of the two objects the time they appear after they have come into contact. By doing so, we eliminated any difference in motion trajectory between the two conditions, while we still maintained a strong impression of streaming or bouncing in the two condition. To assess whether the bounce and streaming events were discriminable with such an intermittent display, we asked an independent group of 8 participants (the author MF and 7 naïve subjects) to perform a simple discrimination task involving bouncing and streaming events. Namely, participants were shown examples of bouncing and streaming trials from Experiment 2 (without the sound), and asked to identify which event was presented (stream vs. bounce). Each participant performed a total of 100 trials (50 streaming events and 50 bouncing events, randomly intermixed). The results show that participants could correctly discriminate stream versus bounce events, on average (± SEM) 92.5% of the times (± 2.2%). Hit rates across the group were also significantly different from chance level (chance = 50%; one-sample t-test, $t(7) = 19.5$, $p < 0.001$), suggesting that despite the intermittent display the nature of the central event was clearly discriminable. In the main experiment, most of the participants completed six blocks of 44 trials, with the exception of one participant who completed five blocks of 44 trials due to the time constrains of the experimental session. Besides the parameters mentioned here, all the other parameters and data analysis procedures were identical to Experiment 1.

Moreover, in Experiment 2, to assess the audio-visual timing range giving rise to a causal impression, we also tested perceived causality. At the end of the main condition, all the participants performed an additional block of 44 trials (in which only the bouncing condition was presented), and were asked to judge whether they perceive a causal relation between the visual and auditory signals (i.e., whether the visual collision appeared to cause the sound), or whether the two events seemed unrelated (i.e., the sound does not originate from the collision). In this condition, we used the same range of audio-visual asynchronies as in the main experiment. Data was analyzed calculating for each participant a point of subjective causality, taken as the mean of best-fitting Gaussian function applied to the proportion of "causal" responses, and with the tolerance of the perceived causality represented by the sigma of the function.

Finally, we also assessed to what extent the difference in visual stimuli between the bounce and the streaming condition can modulate the appearance of causality. In an additional experiment based on the judgment of causality of stream and bounce events we kept all the stimulation parameters identical to the previous causality task, with the exception that both stream and bounce events were intermixed within the same blocks of trials. From participant's responses we calculated the proportion of "causal" judgments, separately for the stream and bounce condition at each SOA. The proportion of "causal" judgments across participants was tested using a two-way repeated measure ANOVA (with factors "condition" and "SOA"), followed by a series of post-hoc paired t-tests within each SOA level. The significance of these post-hoc tests was corrected using the False Discovery Rate (FDR) procedure (Benjamini & Hochberg, 1995) with q = 0.025. Participants performed two blocks of 110 trials.

3.2. Results

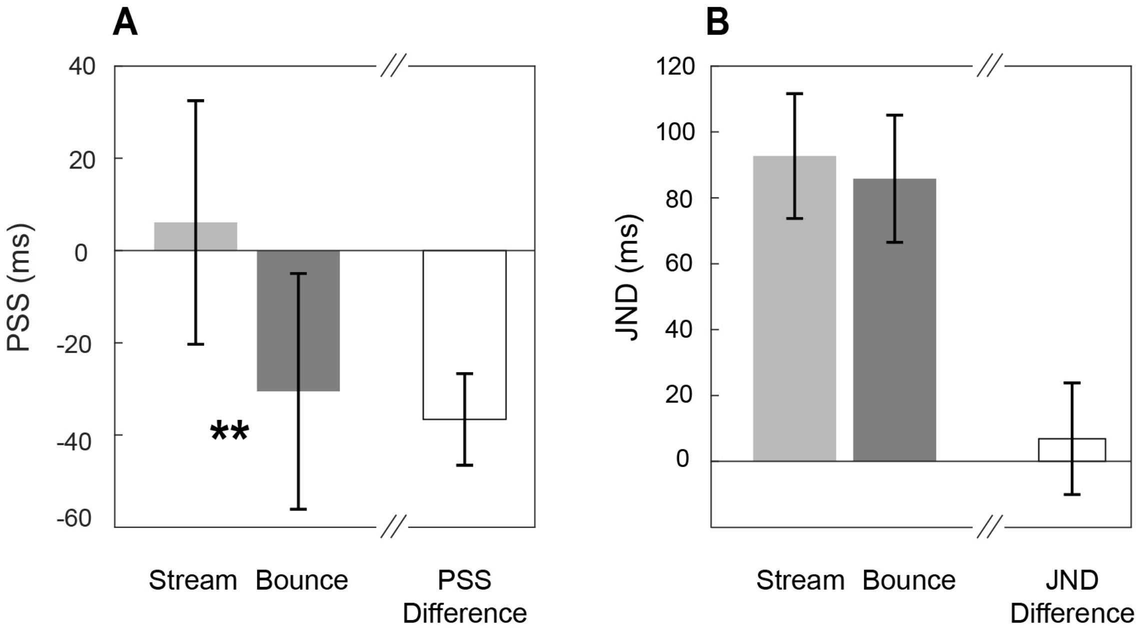


FIGURE 4 – Average PSSs and JNDs in Experiment 2. *(A) PSS results. Changing the spatiotemporal dynamics of the stimuli in order to eliminate any difference between the two conditions, we found again a significant difference between PSS in the streaming and bouncing condition, in the same direction as Experiment 1 (more negative PSS in the bouncing condition). (B) JND results. Also in this experiment, we did not find any significant modulation of JND. The white bar in the rightmost part of both panels indicate the average individual difference between PSS or JND in the two conditions. Error bars are S.E.M.* ** *$p < 0.01$.*

Results obtained with the different visual sequences used in Exp. 2 are reported in Figure 4A and 4B. The pattern of results shows that even though we made the two sequences as similar as possible, we still observed a systematic difference in the average PSS (Figure 4A; 10 ± 28 ms in the stream condition versus -22 ± 26 ms in the bounce condition), with the point of subjective simultaneity in the bounce condition reached with significantly more negative asynchronies (paired sample t-test, bounce versus stream condition; t(12) = 3.68, p = 0.003, *d* = 1.0). This difference indicates that in the bounce condition the PSS was reached with audio-leading pairs, whereas in the stream condition simultaneity was more veridical and obtained with slight visual-leading asynchronies. The average individual difference in PSSs across the two conditions was -37 ± 10 ms. On the other hand, we did not observe any systematic change in JND (93 ± 28 ms, and 90 ± 20 ms, respectively for the stream and bounce condition), as in the previous experiment (t(12) = 0.405, p = 0.69). The average difference in JND across the two conditions was 7 ± 17 ms.

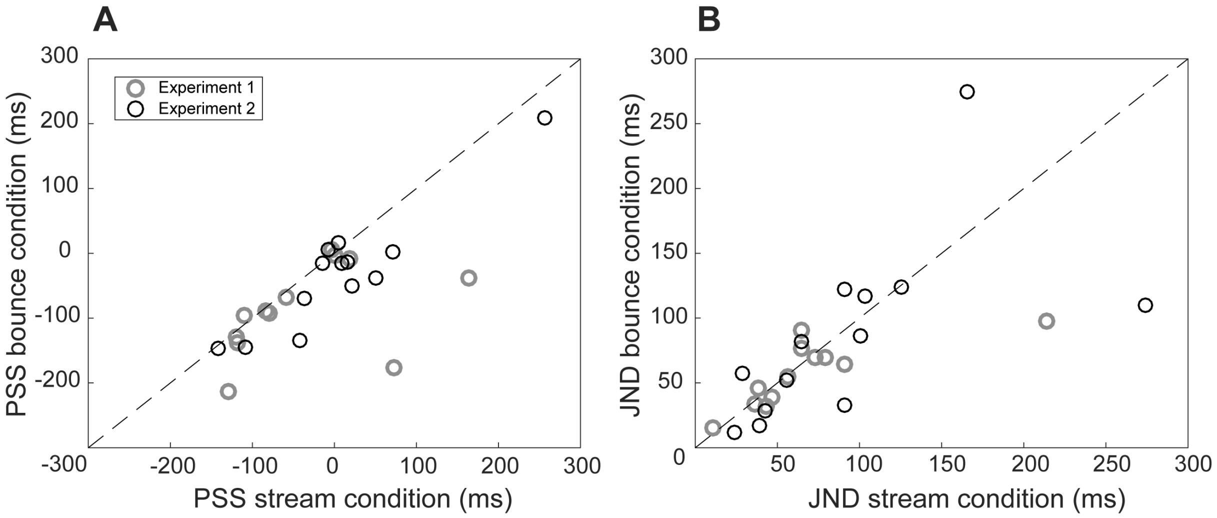


FIGURE 5 – Comparison between individual performances in the stream and bounce condition. *(A) PSS in the bounce condition (y-axis) plotted against the PSS in the stream condition (x-axis). Each dot represents data from a single participant in Exp. 1 (gray bold circles) or Exp. 2 (black circles). Overall, PSSs across the participants are generally lower in the bounce condition compared to the*

stream condition. (B) JND in the bounce condition (y-axis) plotted against the JND in the stream condition (x-axis). Differently from PSSs, JNDs show a much more uniform distribution around the diagonal, showing that none of the conditions systematically affected the level of precision in the TOJ task. The diagonal dashed line indicates the identity line, where PSS or JND in one condition is identical to the other condition.

In order to get a better idea of the performance across the group of participants and across the two experiments, and to ensure that the results are not driven by a few outlier data points, we plotted the individual data (PSS and JND) in the bounce condition against the data in the stream condition, for each participant in the two experiments (Fig. 5). Fig. 5A reports PSSs in the two conditions. Regarding Experiment 1 (gray bold circles), PSSs in the bounce condition appear to be systematically lower than those obtained in the stream condition, for the majority of participants. The average shown in Figure 2, however, may be pushed to lower values due to a few participant (3 out of 12) showing particularly negative PSSs in the bounce condition. Data from Experiment 2 (black circles), instead, show more consistent differences between the two conditions across the group. Overall, Figure 5 shows that PSSs in the bounce condition are systematically lower (more negative, or in some cases less positive) for the majority of participants in the two experiments, demonstrating that the effect is robust and consistent across the group. Figure 5B, on the other hand, reports JND measures. Contrary to the PSS data, JNDs appear to be much more uniformly distributed around the identity line (i.e., the diagonal dashed line), showing that there is little or no systematic effect of condition in both experiments.

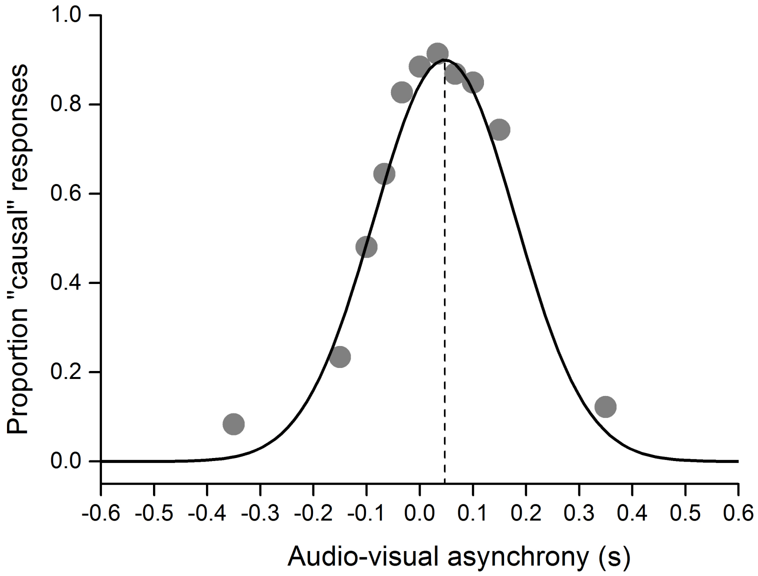


FIGURE 6 – Average distribution of causality judgments. *Results of the causality judgment task, performed in Experiment 2. Gray circles indicate the proportion of "causal" responses as a function of the different levels of audio-visual asynchrony, averaged across participants. The black curve represents the average best-fitting Gaussian function defined by mean and sigma equal to the average Gaussian parameters calculated for each participant.*

Regarding perceived causality, Figure 6 shows the average distribution of causality judgments. The results indicate a clear modulation of perceived causality as a function of the audio-visual asynchrony, and confirmed that participants reliably perceived a cause-effect relation between the auditory and visual events (i.e., the sound was caused by the collision), especially when small asynchronies were presented. On average, perceived causality was reached at 30.2 ± 11.4 ms, so systematically shifted towards visual-first SOAs. Such a positive point of subjective causality resulted to be significantly shifted compared to the physical synchrony of the two events (one-sample t-test, t(12) = 2.76, p = 0.017), again consistently with what might be expected considering the usual temporal order of multisensory events in the real environment (i.e., sound following the visual event causing it). The tolerance of causality perception (i.e., the sigma of the curve) resulted to be relatively large, and quite consistent across participants (102.6 ± 4.4 ms).

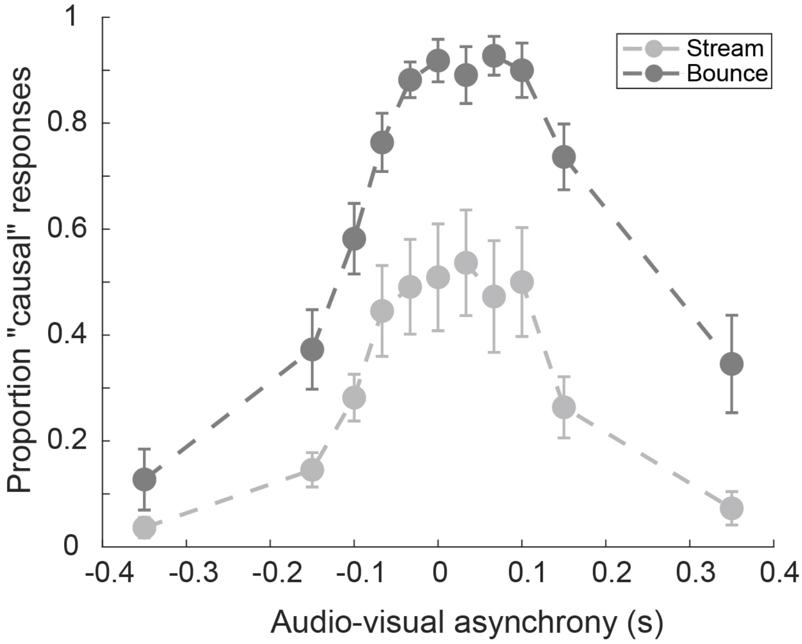


FIGURE 7 – Results of the causality perception experiment including both stream and bounce events. *The plot shows the proportion of "causal" responses as a function of audiovisual SOA, separately for the stream and bounce condition. Negative SOAs indicate audio-first audiovisual pairs. Error bars are SEM.*

Finally, while the causality task performed after the main condition of Experiment 2 provides a measure of the timing of maximum perceived causality of bouncing stimuli, it does not show whether a bounce event facilitates more the perception of causality compared to a stream event. In other words, while our assumption is that only a bounce should lead to apparent causality, it is still possible that even stream events would induce a similar causal impression. To control for this issue, we performed an additional experiment on an independent group of participants, identical to the causality task described above but including both stream and bounce events. Figure 7 shows the results of this control experiment. As shown in the figure, we observed a much higher rate of "causal" responses in the bounce condition, especially at audiovisual SOAs close to physical simultaneity. We performed a two-way repeated measure ANOVA with factor "condition" (stream versus bounce) and "SOA" (the different levels of SOA spanning from -350 to 350 ms around physical simultaneity) on the proportion of responses. The results show a main effect of both condition ($F(1,10) = 17.12$, $p = 0.002$; $\eta_p^2 = 0.63$) and SOA ($F(10,100) = 26.10$, $p < 0.001$; $\eta_p^2 = 0.72$), and a significant interaction between the two factors ($F(10,100) = 4.22$, $p < 0.001$; $\eta_p^2 = 0.30$). Due to this significant interaction, we performed a series of post-hoc tests (paired t-tests) comparing the proportion of "causal" responses in the two conditions separately for each SOA. To account for multiple comparisons, we applied a FDR procedure (Benjamini & Hochberg, 1995) resulting in a critical p-value for significance of 0.01. The results show that the proportion of "causal" responses is significantly higher in the bounce condition compared to the stream condition at most of the SOAs tested. Namely, we observed a significant difference at SOAs spanning from -100 ms to 150 ms ($t(10)$ ranging from 3.16 to 6.29, all p-values < 0.009). No difference was instead observed at the extreme asynchronies (±350ms) and at SOA of -150 ms ($t(10)$ ranging from 1.57 to 2.59, p-values > 0.027).

4. GENERAL DISCUSSION

In this study, we tested whether perceiving a causal link between the auditory and visual signals produced by an event could influence judgments about their temporal order. Causality is indeed pervasive in the external environment, and thus it could potentially represent a powerful prior for the processing of sensory information. To test for this hypothesis, we asked participants to judge the order of the stimuli of an unambiguous stream-bounce display. That is, the display clearly differentiated streaming and bouncing conditions through the spatio-temporal dynamics, color, and shape of the moving objects. Particularly, while in the stream condition the information suggested that the sensory stimuli were not causally related (i.e., the movement of the objects streaming through each other did not suggest a cause for the sound that was heard), the bounce condition suggested that there was a cause-effect relation between the sensory signals (i.e., the collision of two objects suggested a cause for the sound). Using such an arrangement allowed us to exclude the role of visual predictive information (i.e., it cannot be determined whether the event will be a stream or bounce before the two objects reach each other), that while in some studies appeared to be not a crucial factor affecting perceived simultaneity (Kohlraush et al., 2013), it might potentially facilitate temporal order judgments in specific situations. In addition, we extend the validity of our causality manipulation by simulating a bounce between the two objects by swapping the identities of the stimuli only after the visual contact occurs, in order to eliminate any difference in the spatio-temporal dynamics of the two objects. Note that in the following discussion, we sometimes refer to perceptual simultaneity to describe the point of subjective simultaneity (PSS) obtained with a TOJ task, and not to perceived simultaneity *per se* as measured with a SJ task. Previous studies (van Eijk et al., 2008) indeed show that PSS measures obtained with TOJ and SJ tasks do not necessarily correspond, and the SJ task is usually more indicated to assess the perceived simultaneity of audio-visual events *per se*. Instead, a TOJ task as the one used in the present study is more suited to assess an effect on perceived temporal order.

In line with our prediction based on the modulatory effect of expectation and prior knowledge on perception, our results show that the perceived temporal order of audio-visual stimuli is significantly affected by the perception of causality. Particularly, when the objects appear to bounce and change their movement direction – an event that coupled with small audio-visual asynchronies gives the impression that the sound has been generated by the collision – the point of subjective simultaneity obtained with the TOJ task is reached with larger audio-leading SOAs as if there is an expectation that the auditory stimulus should come after the visual event and audio stimuli are perceptually delayed. In fact, when the order of the two stimuli could be better discriminated because no apparent cause-effect relation is suggested, the point of subjective simultaneity (PSS) is closer to veridical simultaneity.

Although we find some differences in the absolute PSS across the two experiments, the biasing effect of bouncing stimuli is consistent in the two cases, suggesting that causality has a robust influence. Results from the causality judgment task also show that perceived causality is maximum when the auditory stimulus follows the visual one by about 30 ms, which is in line with the results of the two TOJ experiments. Taking together the results of the TOJ and causality task, our findings support the idea that perceiving a cause-effect relation between two otherwise unrelated events biases their perceived temporal order. Such a bias is a clear reflection of the usual dynamics of events in the real world, which is constrained by the laws of physics. In other words, our results thus show that causality information is systematically incorporated into the representation of audio-visual timing, consistently with the idea that prior knowledge shapes our perception (e.g., see de Lange et al., 2015 for a review).

First, these results are in line with the literature showing that perceived simultaneity has a much larger tolerance for audio-lagging audio-visual stimuli than for visual-lagging stimuli (e.g., Dixon & Spitz, 1980; Munhall et al., 1996; Grant et al., 2004; Lewald & Guski, 2003; Conrey & Pisoni, 2006). That is, in general, audio lags are more difficult to notice than visual lags. Such a tendency to notice more audio-leading asynchronies is exacerbated by semantic interpretation, not only by low-level stimulus characteristics (Maier et al., 2011). Perceived simultaneity criteria are also been shown to change due to displays that induce the perception of different implied cause-effect relations (van Eijk et al., 2010; Kohlraush et al., 2013), and PSSs obtained with a TOJ task could be affected by these changes in sensitivity (van Eijk et al., 2010) – although previous results were often confounded by the presence of different predictive/postdictive information or by the use of different tasks less optimal for assessing temporal order *per se*.

Nevertheless, our results are consistent with these findings, and show that implied causality between audio-visual stimuli can not only change the tolerance of simultaneity judgments, but can also genuinely affect the perceived temporal order of the events, as measured with a TOJ task. More specifically, the pattern of results observed in the present study is in line with similar trends observed in previous studies. Indeed, van Eijk et al. (2008), employing a TOJ task and comparing stimulus pairs with different complexity (inducing or not the impression of causality), observed less positive PSS measures for stimuli entailing an apparent cause-effect relation between the visual and auditory signal (i.e., bouncing balls). Such an effect could be interpreted as a relative perceptual delay of the auditory signal, which would thus need to be presented slightly earlier compared to audio-visual pairs not providing causal cues (i.e., flash-click pairs). However, the effect in this context was very small, and not statistically significant. A larger effect was instead observed in a subsequent study from the same group (van Eijk et al., 2010), where different causal interpretations were provided by covering one side or the other of a simulated Newton's cradle toy. When the causal interpretation implied a visual event as the cause of a sound, they observed a strong PSS shift towards audio-leading asynchronies. Although in this case the effect was confounded by the presence of predictive information, our results are clearly in line with this finding. Furthermore, the results are also in line with Kohlrausch et al. (2013), providing evidence linking the TOJ PSS shifts observed in van Eijk et al. (2010) with the presence of implied causality between audio-visual stimuli. Namely, they show that a steeper audio-first synchrony boundary – which has been shown to be accompanied by a shift of the TOJ PSS toward more audio-leading asynchronies – most likely reflects the presence of an apparent causal relation between the stimuli rather than predictive information. Our results provide further evidence beyond these previous studies by more clearly showing that apparent causality can directly affect the perceived temporal order of audio-visual stimuli.

On the other hand, we did not observe any effect on the *precision* of perceptual judgments (JND) within each experiment. This shows that the effect of causality, in this context, is limited to the accuracy of perceptual judgments, with little influence on the overall level of performance. Moreover, the modest difference in JND across stream and bounce conditions appears very similar across the two experiments, further suggesting that also the presentation mode (i.e., continuous vs. intermittent) did not affect the JND results. Nevertheless, JNDs appear generally higher in Experiment 2, potentially suggesting that the intermittent presentation may have made the task more difficult. Our experiments, however, were not designed to assess the effect on JNDs of different stimulus dynamics, and the relatively small sample size included in the study did not allow us to assess a between-subject effect with enough power, so that we could not draw strong conclusions from this comparison. However, assessing the effect on JNDs was not the primary aim of this study, and the modest difference in JND between stream and bounce conditions suggests that our experimental manipulation did not provide substantial effects on judgment precision.

A first question, then, is: what is the underlying mechanism giving rise to this bias? An explanation based on purely low-level sensory processing seems unlikely, as in Experiment 2 the visual event in the streaming and bouncing condition was in fact identical until the contact happened. A more likely candidate to explain these results is instead a bias at the perceptual inference level, interpreting sensory information as a function of expectations and previous experience (Grove et al., 2016). For instance, previous studies show that such an inference mechanism may be involved in the stream-bounce illusion (e.g., Sekuler et al., 1997; Fujisaki et al., 2004), where an ambiguous display similar to the one used in the present study is employed. In the stream-bounce illusion, two identical visual stimuli moving towards each other, overlapping, and then continuing their motion, are perceived as streaming upon each other most of the times (~70%; Sekuler et al., 1997; Fujisaki et al., 2004), while they are perceived as colliding and bouncing less often. However, when a sharp sound is introduced, the bouncing interpretation of the ambiguous display becomes more frequent than the streaming one (Fujisaki et al., 2004; Grove & Sakurai, 2009). This phenomenon has been interpreted in several ways, involving early sensory processing, attention, or higher-order cognitive processes. Recently, Grove et al. (2016)

pointed out that a more parsimonious explanation for this effect is a cognitive inference mechanism resolving the ambiguity of the display according to expectations based on previous experience (see also Maier et al., 2011).

In light of this latter point, we propose that a similar mechanism may be involved in the effect of causality observed in the present study. Namely, an inference mechanism may be involved in post-dating the occurrence of the auditory event when its timing is in conflict with the common experience of audio-lagging stimulation. However, differently from Grove et al. (2016), we interpret this effect in terms of *perceptual*, rather than cognitive, inference – that is, in terms of a bias at the high-level perceptual inference stage forming a representation after sensory processing. This idea is in line with Bayesian frameworks of perception (e.g., Knill & Richards, 1996; Mamassian et al., 2002) where previous experience is an important building block of perception. Recent accounts describe a possible computational mechanism for the distortion in perceived stimulus timing by appealing to the combination of a-priori and sensory probabilities of encountering a stimulus (e.g., see Di Luca & Rhodes, 2016). Since humans are frequently exposed to audio-lagging signals in causally-dependent pairs, they may have internalized the probability distribution over asynchronies which exerts a powerful influence able to even distort the perceived temporal order of the stimuli. In this scenario, the observed change in perceived simultaneity would represent a genuine distortion of the perceptual representation of the stimuli, rather than a decision bias. This is indeed also suggested by a recent study investigating the auditory-induced bouncing effect (Meyerhoff & Scholl, 2018). Although this is a different phenomenon, the auditory-induced bouncing effect similarly entails a change in the interpretation of the stimuli due to an apparent causal relation, which has been ascribed to perceptual processing rather than higher cognitive processes. Alternatively, the results may be driven by an explicit response bias, like a tendency of participants to respond “visual first” when they interpreted the bounce as causally determining the sound. This interpretation is however less likely for two reasons. On the one hand, the subtle nature of the bias supports the idea of a perceptual effect, as systematically categorizing audio-visual pairs falling into the causality range as "visual first" would have likely led to a larger bias. On the other hand, if the additional causality information was explicitly used to perform the task, we would expect a difference in precision (i.e., JND) between the two conditions. Indeed, besides potentially increasing the effect, a categorization based on perceived causality in addition (or in alternative) to timing could potentially make responses more consistent, especially at small audio-visual SOAs, thus increasing the overall precision in the task. These arguments remain however speculative at this point, as our behavioral data do not allow to conclusively pinpoint the nature of the effect.

In the context of the effect of expectation on perception, our results are in line with previous studies showing that prior knowledge effectively modulates perceptual processing, altering the appearance of a stimulus (e.g., see de Lange et al., 2018 for a review). Our findings thus expand this literature showing that, first, apparent causality can bias the perceived temporal order of two events, shifting their perceived relative timing according to what would be expected based on the properties of the external environment. Second, our findings suggest that such a mechanism modulating perception operates in a fast and flexible fashion, biasing the stimuli rapidly after the event has been detected. Indeed, participants were not aware in advance of whether the stimuli would stream or bounce. This in turn suggests that the processes mediating the bias most likely unfold right after the detection of the events, flexibly adapting the temporal order representation according to the nature of the interaction. In other words, whether the bias provided by the causality prior emerges or not can be rapidly and flexibly determined directly at the time of perception.

As mentioned above, however, an important point to consider when interpreting the present results is whether participants were actually perceiving a cause-effect relation between the two stimuli. Indeed, we did not explicitly asked participants to report whether they could discriminate bouncing and streaming events, and to what extent the two conditions lead to an impression of causality, during the TOJ task, but only in separate experiments without temporal judgments. Nevertheless, such an effect does not need to be necessarily driven by the conscious perception of a cause-effect relation. Namely, in the presence of a stimulus dynamics consistent with causally related events, the perceptual representation of event timing might be automatically recalibrated even if a participant does not explicitly experience a cause-effect link between the two events. In vision, indeed, several processes operate in the absence of subjective awareness. One example is priming (e.g., see for instance Wiggs & Martin, 1998), whereby a previously unseen stimulus affects the behavioral response to a successive one. Another example is "fast" perceptual adaptation, providing a repulsive aftereffect even when the adaptor stimulus is not discriminable (Glasser et al., 2011) or even masked and not consciously perceived (Fornaciai et al., 2019). Whether or not such a recalibration induced by apparent causality requires awareness of the relation between two stimuli remains however an interesting open question for future studies. Finally, another possibility is that participants may have experienced an impression of causality even in the stream condition. However, data from the control causality task show that although stream events provide some impression of causality, the bounce events produce a stronger and more consistent impression of a cause-effect relation between the visual and the auditory event. Due to the different nature of this control task (i.e., causality judgment, as opposed to the TOJ task) requiring to explicitly attend and judge causality, caution is in order when generalizing these results to the main experiment. Nevertheless, the observation that bounce events more strongly lead to apparent causality is in line with our hypothesis, and supports our conclusion that the shift in perceived temporal order is likely driven by the causality induced by bounce events.

Another question concerns the difference in average PSSs across the two experiments. Indeed, in Experiment 1, PSSs seem to be overall shifted to more negative (audio-leading) values, whereas in

Experiment 2 they are around veridical simultaneity. This difference may have been caused by the difference in the display employed. It is indeed possible that in Experiment 1 the more realistic stream-bounce display showing two objects smoothly moving towards each other may have primed the participants to expect a bounce, shifting the judgments to negative PSSs even in the case of a streaming event, and with an increased effect in the actual bouncing condition. The discontinuous display in Experiment 2, on the other hand, may have lacked such implicit expectation, leading to overall more veridical temporal order estimates. However, our results cannot pinpoint the precise mechanism causing such different distributions of PSS. Nevertheless, as the primary goal of the study was to assess the effect of causality provided by the bounce event, this difference in PSS does not affect our conclusions.

Finally, another interesting point to consider is whether the perceptual delay observed in the present study relates to the delay between light and sound that we would expect between causally-related signals in the real world. We should note that the distance between the participant position and the monitor/loudspeaker was relatively small during the experiment (60 cm) compared to normal observation distances. In the experiment, there should be only a very small delay (1-2 ms) between light coming from the monitor (arriving first) and sound coming from the loudspeaker (arriving later). In contrast, we observe a more substantial perceptual delay in the bounce condition, with a difference of 30-50 ms. This delay is much closer to the expected physical delay of causally-related events that we would find in real-life situations. Considering the speed of sound, the results point to events that should be happening between 10 m and 15 m from the observer. We can thus speculate that although perceptual timing may not necessarily correspond to physical timing, apparent causality may have triggered an automatic expectation of physical delay – a temporal prior that exerts an influence on perception (i.e., see Ernst & Di Luca, 2011, Di Luca & Rhodes, 2016).

Related to this topic, it is worth considering previous studies addressing the problem of perceptual compensation/integration of signals arriving at different times to the respective sensory organs, with such differences in arrival time increasing as a function of distance from the signal source (Sugita & Suzuki, 2003; Kopinska & Harris, 2004; Arnold et al., 2005). For instance, Arnold et al. (2005), addressed this question by exploiting the stream-bounce illusion (and other stimulus displays, in different experiments), and presenting the stimuli from varying source-subject distances. Results from Arnold et al.'s (2005) study show systematic timing shifts varying as a function of distance of the subject from the visual/auditory source. Interestingly, these timing shifts approximated the physical delay between light and sound, suggesting that different sensory signals are not perceptually compensated as a function of their distance (see also Heron et al., 2007 for a similar conclusion). While human participants in previous studies seemed unable to use distance cues to recalibrate perceptual synchrony/temporal order (Arnold et al., 2005; Heron et al., 2007), stronger causality cues like in our unambiguous stream-bounce display may more easily trigger a perceptual recalibration. Investigating the relation between causality, source distance (Arnold et al., 2005) and adaptation (Heron et al., 2007) thus represents another interesting open question for future studies.

5. CONCLUSION

We show that perceived causality can bias perceived temporal order of otherwise unrelated audio-visual stimuli. Namely, the auditory signal tends to be perceived after the visual event that appears to have caused it, even if the audio was actually presented up to 40 ms before the visual event. This finding shows that the perceived cause-effect relationship suggested by an event in the real world can provide a powerful expectation about the timing of the associated signals which is able to distort the perceived temporal order. These results provide further evidence concerning the active and constructive nature of perception, whereby an inference mechanism creates perception not only based on sensory signals, but also on previous knowledge and expectations.

AUTHOR CONTRIBUTIONS

MDL and MF conceived and designed the study; MF collected and analyzed the data; MDL and MF interpreted the results and wrote the manuscript.

ACKNOWLEDGMENTS

This research was funded by Marie Curie CIG 304235 ‘TICS’.

COMPETING INTERESTS

The Authors declare no competing interests.

DATA AVAILABILITY

All the data generated during the experiments described in this manuscript is available on Open Science Framework at the this link: https://osf.io/79rty/

REFERENCES

Allison, T., Wood, C. C., & Goff, W. R. (1983). Brain stem auditory, pattern-reversal visual, and short-latency somatosensory evoked potentials: Latencies in relation to age, sex, and brain and body size. *Electroencephalography and Clinical Neurophysiology*, *55*(6), 619–636. http://doi.org/10.1016/0013-4694(83)90272-9

Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*, *45*(10), 1275–1284. http://doi.org/10.1016/j.visres.2004.11.014

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, *119*(6), 4065–4073. http://doi.org/10.1121/1.2195091

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? Trends in Cognitive Sciences, 22(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Di Luca, M., & Rhodes, D. (2016). Optimal Perceived Timing: Integrating Sensory Information with Dynamically Updated Expectations. *Scientific Reports*, *6*, 28563. http://doi.org/10.1038/srep28563

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*. http://doi.org/10.1068/p090719

Ernst, M. O., & Luca, M. Di. (2011). Multisensory Perception: From Integration to Remapping. In *Sensory Cue Integration* (pp. 224–250). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.003.0012

Fornaciai, M., & Park, J. (2019). Spontaneous repulsive adaptation in the absence of attractive serial dependence. *Journal of Vision*, 19(5), 21. https://doi.org/10.1167/19.5.21

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. http://doi.org/10.1038/nn1268

Glasser, D. M., Tsui, J. M. G., Pack, C. C., & Tadin, D. (2011). Perceptual and neural consequences of rapid motion adaptation. *Proceedings of the National Academy of Sciences*, 108(45), E1080–E1088. https://doi.org/10.1073/pnas.1101141108

Grant, K. W., Wassenhove, V. van, & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication*, *44*(1–4), 43–53. http://doi.org/10.1016/j.specom.2004.06.004

Grove, P. M., Robertson, C., & Harris, L. R. (2016). Disambiguating the Stream/Bounce Illusion With Inference. *Multisensory Research*, *29*(4–5), 453–464. http://doi.org/10.1163/22134808-00002524

Grove, P. M., & Sakurai, K. (2009). Auditory Induced Bounce Perception Persists as the Probability of a Motion Reversal is Reduced. *Perception*, *38*(7), 951–965. http://doi.org/10.1068/p5860

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, *7*(13), 5. http://doi.org/10.1167/7.13.5

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception ECVP 2007 Abstract Supplement*, *36*(14), 1–16. http://doi.org/10.1068/v070821

Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambrige: Cambridge University Press.

Kohlrausch, A., van Eijk, R., Juola, J. F., Brandt, I., & van de Par, S. (2013). Apparent causality affects perceived simultaneity. *Attention, Perception & Psychophysics*, *75*(7), 1366–73. http://doi.org/10.3758/s13414-013-0531-0

Kopinska, A., & Harris, L. R. (2004). Simultaneity Constancy. *Perception*, *33*(9), 1049–1060. http://doi.org/10.1068/p5169

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468–478. http://doi.org/10.1016/S0926-6410(03)00074-0

Maier, J. X., Di Luca, M., & Noppeney, U. (2011). Audiovisual Asynchrony Detection in Human Speech. *Journal of Experimental Psychology: Human Perception and Performance*. http://doi.org/10.1037/a0019952

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic Models of the Brain; Perception and Neural Function.* (pp. 13–36). Cambridge, MA: MIT Press.

Matteson, H. H. (1971). Effects of Surround Luminance on Perceptual Latency in the Fovea. *Journal of the Optical Society of America*, *61*(9), 1169. http://doi.org/10.1364/JOSA.61.001169

Meyerhoff, H. S., & Scholl, B. J. (2018). Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents. *Cognition*, *170*, 88–94. http://doi.org/10.1016/j.cognition.2017.08.007

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics*. http://doi.org/10.3758/BF03206811

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. Journal of Vision, 15(8), 13. https://doi.org/10.1167/15.8.13

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

Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. Current Biology, 21(21), R883–R884. https://doi.org/10.1016/j.cub.2011.09.013

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. Journal of Experimental Psychology: General, 144(6), 1089–1104. https://doi.org/10.1037/xge0000109

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. Current Biology, 18(16), R697–R698. https://doi.org/10.1016/j.cub.2008.06.021

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature*, *421*(6926), 911. http://doi.org/10.1038/421911a

Van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & Van De Par, S. (2010). Temporal order judgment criteria are affected by synchrony judgment sensitivity. *Attention, Perception, & Psychophysics*, *72*(8), 2227–2235. http://doi.org/10.3758/BF03196697 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. http://doi.org/10.3758/PP.70.6.955

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*(5), 515–522. http://doi.org/10.1016/0042-6989(79)90136-6

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. Current Opinion in Neurobiology, 8(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X