Temporal dynamics of a perceptual decision

Mick Zeljko
School of Psychology, The University of Queensland, Brisbane, Australia

Ada Kritikos
School of Psychology, The University of Queensland, Brisbane, Australia

Philip M. Grove
School of Psychology, The University of Queensland, Brisbane, Australia

Previous research suggests that cognitive factors acting in a top-down manner influence the perceptual interpretation of ambiguous stimuli. To examine the temporal unfolding of these influences as a perceptual decision evolves, we have implemented a modified version of the stream-bounce display. Our novel approach allows us to track responses to stream-bounce stimuli dynamically over the entire course of the motion sequence rather than collecting a subjective report after the fact. Using a trackpad, we had participants control a cursor to track a stream-bounce target actively from start to end and measured tracking speed throughout as the dependent variable. Our paradigm replicated the typical effect of visual-only displays being associated with a streaming bias and audiovisual displays with a bouncing bias. Our main finding is a significant behavioral change preceding a perceptual decision that then predicts that decision. Specifically, for trials in which the sound was presented, tracking speeds were significantly slower starting 500 ms before the point of coincidence and presentation of the sound for bounce compared to stream responses. We suggest that behavioral response may reflect a cognitive expectation of a perceptual outcome that then biases action and the interpretation of sensory input to favor that forthcoming percept in a manner consistent with both the predictive-coding and common-coding theoretical frameworks. Our approach provides a novel behavioral corroboration of recent imaging studies that are suggestive of early brain activity in perception and action.

Introduction

Evidence suggests that two complementary sources of information about the world contribute to the formation of subjective perceptual experience: sensory input and prior knowledge. Sensory input provides information that reflects the current state of the world, but it is limited and therefore always (varyingly) ambiguous. Prior knowledge, gained through experience, provides information about how the world works (e.g., Gekas, Seitz, & Seriès, 2015; Gilbert & Sigman, 2007; Kersten, Mamassian, & Yuille, 2004; Kornmeier, Hein, & Bach, 2009; Maloney, Dal Martello, Sahm, & Spillmann, 2005; Summerfield & Egner, 2009; Wang, Arteaga, & He, 2013). To construct a subjective percept that accurately reflects the current state of the world, the brain uses both information sources and makes “the best sense of sensory inputs based on a set of hypotheses or constraints derived by prior knowledge and contextual influences” (Gilbert & Sigman, 2007, p. 678).

Recent work has considered the influence of prior knowledge in the perception of ambiguous multisensory stimuli using the stream-bounce display (an ambiguous motion sequence in which two identical targets moving along intersecting trajectories are typically seen to either stream past or bounce off one another). Typically, the presence or absence of a brief sound at the point of coincidence of the targets biases responses: Sounds are reliably associated with increased bounce reports (Sekuler, Sekuler, & Lau, 1997; Zeljko & Grove, 2017a, 2017b). While stimulus manipulations at (or close to) the point of coincidence can modulate responses (for example, brief damped sounds lead to increased bouncing compared to brief ramped sounds; Grassi & Casco, 2009), other, seemingly extrastimulus factors also modulate responses. For example, Grove, Robertson, and Harris (2016) examined the influence of expectation using a modified stream-bounce display in which the targets moved on horizontal trajectories that were slightly vertically offset. The offset targets always objectively streamed, so to report a bounce, an observer would need to accept that the targets had undergone a small vertical shift (a switch) at the point of coincidence. The researchers found that introducing trajectory
switches prior to the point of coincidence increased the proportion of reported bounce percepts, with increasing precocincidence switches associated with increasing reported bouncing. They concluded that switches prior to coincidence created expectations that primed perceptual inference to modulate subsequent perceptual decisions.

While the effects of prior knowledge on perception are well established, the specific mechanisms by which it exerts an influence remain uncertain (e.g., Sherman, Kanai, Seth, & VanRullen, 2016). Extrastimulus factors like expectation may exert an early influence, biasing sensory processing to favor a particular outcome, or they may exert a later influence, biasing the interpretation of sensory information during perceptual decision making or response selection. Our aim here is to examine the temporal unfolding of a perceptual decision using the stream-bounce display to determine when extrastimulus factors may be exerting their influence. To do this, we implemented a modified version of the stream-bounce display. Rather than the typical method of collecting a subjective report of streaming or bouncing after the event, we tracked participant responses dynamically over the course of the extended event, collecting rich data sets of high temporal resolution.

Our response-tracking approach follows several recent studies that have used mouse tracking to probe the cognitive processes underlying various choice tasks (e.g., Bonnen, Burge, Yates, Pillow, & Cormack, 2015; Dale, Kehoe, & Spivey, 2007; Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Hehman, Stolier, & Freeman, 2015; Rheem, Verma, & Becker 2018; Spivey, Grosjean, & Knoblich, 2005). In a choice task, a measure is typically made of some function of the choice outcome. For example, the frequency with which each outcome is made may be reported, or the response time taken to make a choice may be measured and then averaged for each outcome. While informative, what these types of measures do not provide is information on the degree to which each alternative outcome received consideration during the choice process and how commitment to and conflict between options evolves (Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, in press). Mouse-tracking studies assume that motor movements in a time interval contain a signal of cognitive processes during that interval (Spivey & Dale, 2006) and so motion trajectories reflect underlying cognitive processes (Hehman et al., 2015). Analyses of these mouse trajectories and their temporal dynamics can therefore provide insights into these underlying processes (Hehman et al., 2015). Response tracking has been used to examine language processing (Spivey et al., 2005), target selection in visual search (Song & Nakayama, 2008), visual sensitivity (Bonnen et al., 2015), and cognitive load (Rheem et al., 2018).

Using a trackpad, we had participants control a cursor to track one stream-bounce target actively throughout the entire motion sequence. By recording the cursor position with high temporal resolution, we determined tracking velocity throughout the motion sequence. By classifying responses as either streaming or bouncing based on the final position of the cursor, our approach enabled us to examine how tracking velocity varies throughout as the stream-bounce event evolves, and to do so separately for stream and bounce responses.

We contend that this approach allows us to track when prior knowledge may be exerting influence on perceptual decisions. We have two hypotheses. First, we predict that tracking velocities will diverge abruptly at some point after the targets have coincided when a response (stream or bounce) is actioned. At this point tracking velocities will clearly diverge, with stream responses showing a continuation of tracking in the same direction and bounce responses showing a reversal of tracking. This postcoincidence divergence reflects the actioning of a response, and we emphasize it to distinguish it from more subtle deviations in velocity that may reflect subtle biases. Our second hypothesis is that if extrastimulus factors influence perceptual decisions, then these factors may manifest as behavioral differences in the lead-up to perceptual decisions that correlate with the different perceptual outcomes. Specifically, if prior knowledge influences perceptual decisions early, then we predict a divergence in tracking velocities for streaming versus bouncing responses prior to the divergence with reversal that indicates the actioning of the response.

## Methods

### Participants

Thirty subjects (24 female, six male; age: 23.17 ± 8.22 years) from the University of Queensland—the main participant group—participated in the study in return for course credit. All participants reported normal hearing and normal or corrected-to-normal vision, and all except the authors were unaware of the purpose of the experiment. All experiments were cleared in accordance with the ethical-review processes of the University of Queensland and within the guidelines of the National Statement on Ethical Conduct in Human Research.

### Apparatus and stimuli

Stimuli were generated on a Mac Mini (2.5-GHz Intel Core i5 processor with 4 GB of 1,600-MHz DDR3
memory and an Intel HD Graphics 4000 1,024-MB graphics chip and running OS X 10.9.5) using MATLAB (R2015b; MathWorks, Natick, MA) and the Psychophysics Toolbox extensions (V3.0.11; Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard 2007). Visual stimuli were viewed on an Apple Thunderbolt Display (resolution: 2,560 × 1,440), sounds were presented via a set of Sony MDR-XB450 headphones, and tracking data were collected via an Apple Magic Trackpad 2 (width: 160 mm). The display was viewed from approximately 80 cm, and the trackpad was positioned directly below the display.

Visual stimuli consisted of either two distinguishable discs (one black and one white on a gray background) in the case of objective motion or two identical discs (black on a gray background) in the case of subjective motion. The discs were separated horizontally about the midline of the display and arranged to allow motion on a path diagonally down (at an angle of 45°) toward the display center, intersecting 1° above a fixation cross placed at the display center (the midpoint of the motion sequence). Each disc was 0.4° in diameter, the path was 16° long, and the fixation cross subtended 0.3°. A red spot (0.3° in diameter) was used to indicate which disc would be the target to track on each trial, and a black rectangular cursor (0.66° × 0.1°) was used to track the target. The cursor was free to move horizontally but was constrained vertically to motion on a horizontal line 5° below the central fixation cross. The dimensions of the motion sequence were selected in order that the horizontal range of motion of the cursor would match the physical range of motion permitted by the trackpad, providing a one-to-one correspondence between cursor and hand motion.

The motion sequence for a single trial lasted for 2 s and consisted of 121 frames presented at a frame rate of 60 Hz. The 121 frames included a central coincidence frame (Frame 61, the point of coincidence [PoC], in which the targets completely overlapped, and 60 frames of motion before and after the PoC. Upon motion initiation, the discs were sequentially displaced toward the display midline on a diagonal/downward trajectory at a constant apparent speed of 8°/s (Frames 1–60), coincided (Frame 61), and then continued their motion (Frames 62–120) away from the display midline to the end point of the sequence (Frame 121).

The auditory stimulus was a 15-ms 800-Hz tone, amplitude modulated with an exponential decay (time constant 5 ms) and sampled at 44.1 kHz. The average sound pressure level of the tone was approximately 65 dB SPL measured at the headphone earpiece (with a Lutron SL-4012 sound level meter). Ambient sound level was approximately 45 dB SPL.

**Procedure**

Participants were seated unrestrained in front of the keyboard, trackpad, and display with their eyes approximately 80 cm from the display, and the essential aspects of the stimulus and motion sequence were described to them. Participants were then instructed that their task was to control the cursor using the trackpad with their dominant hand and to try and keep it directly below the target disc throughout the entire motion sequence. “Staying with the target” was emphasized in contrast to “waiting and following” or “anticipating and leading” the motion, and participants were encouraged to try to look at the fixation cross for the entire motion sequence. A block of trials was initiated by pressing the space bar on the keyboard.

A single trial consisted of the following sequence of events: (1) At the start of the trial, only the fixation cross was visible, for 750 ms; (2) the discs appeared at their start positions and the cursor appeared at its start position (horizontal center, 5° below the display center), and the cursor control was disabled for 500 ms; (3) the red indicator spot appeared 0.4° above one of the targets and the cursor control was enabled; (4) using the trackpad, the participant moved the cursor horizontally until it was directly below the indicated target; (5) when the cursor was directly below the indicated target, the indicator spot vanished and the cursor control was disabled for 250 ms; (6) motion was initiated and the cursor control was enabled; (7) the targets moved for 2,000 ms while the participant attempted to track the target disc with the cursor (for sound trials in the subjective-motion block, a brief tone was played at the PoC); (8) the discs and the cursor stopped and the cursor control was disabled for 750 ms; and (9) the discs and cursor vanished, leaving only the fixation cross (see Figure 1.).

All participants completed four consecutive blocks in the following order: 32 trials of objective motion (a practice block), 80 trials of objective motion (a test block), 16 trials of subjective motion (a practice block), and 80 trials of subjective motion (a test block). For the objective-motion blocks, half of the trials objectively streamed and half objectively bounced, and these were randomly intermixed. Color starting side (left or right), target side (left or right), and disc on top at coincidence (black or white) were counterbalanced within subjects. For subjective motion, half of the trials had a sound presented at PoC and half had no sound (randomized), and target side (left or right) was counterbalanced.

**Data**

Only data from test blocks (both objective and subjective motion) were recorded for analysis. For each
trial, we collected raw tracking data (the x-coordinate of the center of the cursor in the display frame of reference) for each of the 121 frames comprising the motion sequence within the trial. The raw tracking data were then used to determine tracking distance, tracking velocity, tracking-epoch velocity, and tracking response according to the following definitions.

**Tracking distance**

Tracking distance is the magnitude of the horizontal distance of the cursor from its starting position (Frame 1) for each frame, measured in pixels. The tracking distance is positive regardless of starting side and starts at zero. Specifically, the tracking distance should increase throughout the motion sequence for streaming targets (i.e., the final cursor position will be opposite the initial cursor position) and should increase to some maximum then decrease toward zero throughout the motion sequence for bouncing targets (i.e., the final cursor position will be on the same side, and close to, the initial cursor position).

**Tracking velocity**

Tracking velocity is the single-frame change in distance of the cursor for each frame (except Frame 1), measured in pixels per frame. The tracking velocity is positive when the cursor moves away from its initial position and negative when it moves toward its initial
position. Specifically, the tracking velocity should remain positive throughout the motion sequence for streaming targets (i.e., tracking remains unidirectional away from the initial cursor position) and should be initially positive and then negative for bouncing targets (i.e., tracking is initially away from the initial cursor position then reverses toward the initial cursor position).

**Tracking-epoch velocity**

Tracking-epoch velocity is the change in distance of the cursor for each of the eight 250-ms epochs of the motion sequence, measured in pixels per frame. The tracking-epoch velocity represents the average tracking velocity over the epoch and is positive when the cursor moves away from its starting position and negative when it moves toward its starting position.

**Tracking response**

The tracking response is either “stream” or “bounce” depending on whether the final position of the cursor (Frame 121) was on the opposite or the same side of the display midline as the starting position (Frame 1).

**Results**

We first tested to ensure that in the objective-motion block, tracking responses accurately reflect objective streaming and bouncing motion, and that in the subjective-motion block they produce the typical stream-bounce effect of significantly more bounce responses for sound trials compared to no-sound trials. We computed the percentage of trials yielding bounce responses for each participant in each condition of the two motion blocks to determine the group mean percentage of bounce responses (objective motion: streaming and bouncing; subjective motion: sound and no-sound).

Paired t tests of the percentage of bounce responses within each motion block revealed that objective motion tracking was highly accurate—$M(\text{Str}) = 0\%$, $SE(\text{Str}) = 0\%$, $M(\text{Bnc}) = 97\%$, $SE(\text{Bnc}) = 2\%$, $t(29) = 53.92$, $p < 0.001$—and that the typical stream-bounce effect (sound modulation of responses) occurred: $M(\text{No-Snd}) = 6\%$, $SE(\text{No-Snd}) = 2\%$, $M(\text{Snd}) = 61\%$, $SE(\text{Snd}) = 6\%$, $t(29) = 9.33$, $p < 0.001$ (see Figure 2). We conclude that our paradigm is valid in that participants can track a moving target that either streams or bounces, and tracking responses reproduce the standard stream-bounce effect.

We next considered the dynamics of motion tracking by conducting several temporal analyses. For each participant, we first computed mean tracking-distance and tracking-velocity data for each frame of motion, for each tracking-response type (stream or bounce), and in each motion block (objective and subjective) to determine the general characteristics of tracking (see Figure 3). We note that for both objective- and subjective-motion tracking, tracking performance is accurate, as evidenced by the tracking-distance profiles: The cursor position closely matches the target position throughout the motion sequence except immediately following motion onset and reversal. Tracking velocity further reflects this with an initial delayed response to motion initiation, a period of reasonably accurate velocity matching, a delayed response to a streaming or bouncing decision, and finally an overcorrection to counteract the streaming/bouncing delay. Performance across participants and trials in both distance and velocity is highly consistent, as evidenced by the narrow spread of the standard error of the means.

We next conducted planned paired t tests comparing the tracking velocity of stream versus bounce responses within each frame of the motion sequence in each motion block (subjective and objective) to determine when the tracking velocities reliably diverged (refer to...
Figure 4a and 4c). A reliable divergence was defined as a minimum of three consecutive frames that were significant at the $p<0.05$ level for objective motion and a minimum of four consecutive frames for subjective motion. These criteria were established using the method described by Dale et al. (2007), whereby 10,000 simulated experiments using the same participant numbers and frame-by-frame means and standard deviations were used to determine the frequency with which sequences of significant differences occurred (see also Farmer et al., 2007). For objective motion, significantly different sequences of one, two, and three frames occurred by chance 77%, 6%, and 0.4% of the time, respectively, and for subjective motion, significantly different sequences of one, two, three, and four frames occurred by chance 60%, 18%, 7%, and 3% of the time.

For objective-motion tracking, Frame 76 was the first to diverge reliably (i.e., 250 ms after the PoC)—$M(\text{Str}) = 6.2$, $SE(\text{Str}) = 0.2$, $M(\text{Bnc}) = 5.4$, $SE(\text{Bnc}) = 0.3$, $t(29) = 2.33$, $p = 0.028$—and after this point, the tracking velocity for stream responses was significantly different from that for bounce responses for all remaining frames until the end of the motion sequence. For subjective-motion tracking, we found that tracking velocities for stream and bounce responses reliably diverged before the PoC. The first frame in which velocities reliably differ is Frame 53, 133 ms before the PoC: $M(\text{Str}) = 5.6$, $SE(\text{Str}) = 0.2$, $M(\text{Bnc}) = 4.9$, $SE(\text{Bnc}) = 0.2$, $t(27) = 2.62$, $p = 0.014$. We note that Frame 39 was the first to diverge at the $p<0.05$ level (367 ms before the PoC: $M(\text{Str}) = 4.7$, $SE(\text{Str}) = 0.3$, $M(\text{Bnc}) = 4.3$, $SE(\text{Bnc}) = 0.3$, $t(27) = 2.33$, $p = 0.027$). Although Frame 39 does not diverge reliably, it appears to signify the beginning of a systematic divergence in tracking velocities between stream and bounce responses.

As an additional statistical test, we computed pooled bins (tracking velocities averaged within 250-ms epochs) to compare tracking to perceived streaming versus bouncing targets over the entire motion sequence using analysis of variance (e.g., Dale et al., 2007; Farmer et al., 2007). We consider objective-motion tracking first. A two-factor analysis of variance
Figure 4. Group mean tracking velocities for streaming (solid lines) and bouncing (dotted lines) responses for objective motion—(a) frame by frame and (b) in 250-ms epochs ± standard error of the mean—and for subjective motion: (c) frame by frame and (d) in 250-ms epochs ± standard error of the mean. In all panels, the red line indicates the point of coincidence. In (a, c), the straight black lines indicate the target velocity and the black bars indicate frames for which the streaming and bouncing velocities are significantly different (p < 0.05). For clarity, the y-axis in (a, c) has been truncated at 0 and so does not show the negative velocities of the bounce responses.

(ANOVA; response: stream or bounce; epoch: 1–8) revealed significant main effects of response, \( F(1, 29) = 57.835, p < 0.001, \eta_p^2 = 0.999 \), and epoch, \( F(7, 203) = 208, p < 0.001, \eta_p^2 = 0.878 \), and a significant interaction between the two, \( F(7, 203) = 322, p < 0.001, \eta_p^2 = 0.917 \) (see Figure 4b). Planned follow-up comparisons (Bonferroni corrected) revealed that the tracking-epoch velocities for stream and bounce responses were not significantly different until Epoch 6 (i.e., the epoch covering the period 250–500 ms after the PoC) and then for the following two epochs (7 and 8)—E1: \( M(\text{Str}) = 1.45, SE(\text{Str}) = 0.27, M(\text{Bnc}) = 1.28, SE(\text{Bnc}) = 0.25, t(27) = 1.53, p = 0.138 \); E2: \( M(\text{Str}) = 7.96, SE(\text{Str}) = 0.35, M(\text{Bnc}) = 8.05, SE(\text{Bnc}) = 0.34, t(27) = -0.82, p = 0.420 \); E3: \( M(\text{Str}) = 6.61, SE(\text{Str}) = 0.22, M(\text{Bnc}) = 6.70, SE(\text{Bnc}) = 0.24, t(27) = -0.69, p = 0.495 \); E4: \( M(\text{Str}) = 5.68, SE(\text{Str}) = 0.20, M(\text{Bnc}) = 5.58, SE(\text{Bnc}) = 0.19, t(27) = 0.88, p = 0.385 \); E5: \( M(\text{Str}) = 4.82, SE(\text{Str}) = 0.21, M(\text{Bnc}) = 4.76, SE(\text{Bnc}) = 0.22, t(27) = 0.71, p = 0.482 \); E6: \( M(\text{Str}) = 8.11, SE(\text{Str}) = 0.29, M(\text{Bnc}) = -3.84, SE(\text{Bnc}) = 0.69, t(27) = 15.22, p < 0.001 \); E7: \( M(\text{Str}) = 6.44, SE(\text{Str}) = 0.27, M(\text{Bnc}) = -15.90, SE(\text{Bnc}) = 0.73, t(27) = 29.28, p < 0.001 \); E8: \( M(\text{Str}) = 3.75, SE(\text{Str}) = 0.22, M(\text{Bnc}) = -5.97, SE(\text{Bnc}) = 0.39, t(27) = 18.38, p < 0.001 \).

A similar two-factor ANOVA (response: stream or bounce; epoch: 1–8) conducted on the subjective-motion tracking-epoch-velocity data also revealed significant main effects of response, \( F(1, 27) = 23.507, p < 0.001, \eta_p^2 = 0.999 \), and epoch, \( F(7, 189) = 109, p < 0.001, \eta_p^2 = 0.801 \), and a significant interaction between the two, \( F(7, 189) = 197, p < 0.001, \eta_p^2 = 0.879 \) (see Figure 4d). Planned follow-up comparisons (Bonferroni corrected), however, revealed that for subjective motion, tracking-epoch velocities for stream and bounce responses were significantly different for Epochs 4–8 (i.e., for one full 250-ms epoch before the PoC and for all epochs following) and with a clear trend commencing at Epoch 3—E1: \( M(\text{Str}) = 2.57, SE(\text{Str}) = 0.48, M(\text{Bnc}) = 2.70, SE(\text{Bnc}) = 0.51, t(27) = -0.78, p = 0.444 \); E2: \( M(\text{Str}) = 7.62, SE(\text{Str}) = 0.33, M(\text{Bnc}) = 7.43, SE(\text{Bnc}) = 0.46, t(27) = 0.42, p = 0.678 \); E3: \( M(\text{Str}) = 6.27, SE(\text{Str}) = 0.21, M(\text{Bnc}) = 5.73, SE(\text{Bnc}) = 0.16, t(27) = 2.75, p = 0.011 \); E4: \( M(\text{Str}) = 5.61, SE(\text{Str}) = 0.19, M(\text{Bnc}) = 5.00, SE(\text{Bnc}) = 0.12, t(27) = 3.40, p = 0.002 \); E5: \( M(\text{Str}) = 4.87, SE(\text{Str}) = 0.28, M(\text{Bnc}) = 3.08, SE(\text{Bnc}) = 0.33, t(27) = 6.79, p < 0.001 \); E6: \( M(\text{Str}) = 7.55, SE(\text{Str}) = 0.38, M(\text{Bnc}) = -8.54, SE(\text{Bnc}) = 0.81, t(27) = 17.61, p < 0.001 \); E7: \( M(\text{Str}) = 6.12, SE(\text{Str}) = 0.28, M(\text{Bnc}) = -9.79, SE(\text{Bnc}) = 0.47, t(27) = 31.39, p < 0.001 \); E8: \( M(\text{Str}) = 3.95, SE(\text{Str}) = 0.28, M(\text{Bnc}) = -5.11, SE(\text{Bnc}) = 0.42, t(27) = 14.95, p < 0.001 \).
Given our earlier finding that stream and bounce responses show a strong dependency on the presence or absence of the sound, we next considered the sound dependence of the tracking velocity as a control analysis (Figure 5). Repeating the frame-by-frame analysis for stream versus bounce responses, we found the same pattern of results for both sound (Figure 5a) and no-sound trials (Figure 5b) that we found for all trials—that is, an early, precoincidence divergence of tracking velocities that is not seen with objective-motion tracking (we note that for no-sound trials there is less sequential significance than for sound trials, but this is likely due to the low proportion of bounce responses for no-sound trials). As a final sound-dependency check, we compared the frame-by-frame tracking velocity for sound versus no-sound trials (ignoring the response), and as expected found no early effects (in fact, the sound-versus-no-sound tracking-velocity profiles for subjective motion strongly resemble the stream-versus-bounce profiles for objective motion; Figure 5c).

Perceptual decisions regarding ambiguous stimuli have been shown to be biased toward recent decisions, causing a clustering of responses (e.g., Brascamp, Kanai, Walsh, & van Ee, 2010; Maloney et al., 2005; Murphy, Leopold, & Welchman, 2014; Pearson & Brascamp, 2008). To examine potential sequential effects, we conducted a sequential-dependency analysis of both stream and bounce responses to subjective motion and their associated tracking-velocity profiles.

We categorized each trial according to whether the response to the previous trial was a stream or a bounce and according to whether there was a sound, and computed the percentage of trials yielding bounce responses for each participant in each condition. We conducted a two-factor ANOVA (previous response: stream or bounce; sound: present or absent) on bounce responses and found significant main effects of previous response, $F(1, 27) = 7.17, p = 0.012$, $\eta^2_p = 0.210$, and sound, $F(1, 27) = 144, p < 0.001$, $\eta^2_p = 0.842$, and a significant interaction, $F(1, 27) = 10.46, p = 0.003$, $\eta^2_p = 0.279$. Follow-up paired $t$ tests to examine the effect of previous response for each sound condition revealed that there was a significant serial dependency in bounce responses for sound trials—$M(PrevStr) = 62.0\%, SE(PrevStr) = 5.9\%, M(PrevBnc) = 77.7\%, SE(PrevBnc) = 5.5\%, t(28) = -3.01, p = 0.006$—but not for no-sound trials: $M(PrevStr) = 6.4\%, SE(PrevStr) = 2.7\%$.
We next conducted planned paired $t$ tests comparing the tracking velocity of previous stream versus previous bounce responses within each frame of the motion sequence to determine when the tracking velocities reliably diverged (Figure 6b). A reliable divergence was defined as a minimum of three consecutive frames that were significant at the $p < 0.05$ level, again established using the simulated experimental procedure described earlier (for these data, significantly different sequences of one, two, and three frames occurred by chance 75%, 5%, and 0.4% of the time, respectively). While there are two short (four-frame; 67-ms) segments where the previous response results in a reliable velocity difference (one precoincidence commencing at Frame 18 and one postcoincidence commencing at Frame 69), the serial dependence on tracking velocity is small and the temporal characteristics are fundamentally different from those of the current-response dependence on tracking velocity (Figure 4c).

To summarize, using a valid adaptation of the stream-bounce effect we demonstrated through temporal analyses of motion tracking that hand tracking velocity is significantly different for bounce compared with stream responses. This commenced 250 ms before the point of coincidence of the targets and 500 ms earlier than observed for objectively streaming versus bouncing targets. We emphasize that this difference was on a within-observer basis and so was unlikely due to between-observers biases for either streaming or bouncing outcomes. This pattern of results occurred for trials in which there was a sound present. Finally, although there was a serial dependency on bounce responses, it does not appear to account for the tracking-velocity findings. Taken together, these results suggest that the presence of a sound at coincidence, a preparatory slowing of tracking over an extended period prior to coincidence, and a bounce response on the previous trial all positively influence the likelihood of a bounce response. Importantly, though, these three factors appear to exert their influences independently.

**Controls**

Our paradigm differs from the typical stream-bounce design in at least four key aspects, and each of these might obscure the interpretation of our findings. First, our design requires attentional tracking of one of the two discs, while in typical stream-bounce experiments this is explicitly discouraged. Second, it requires active tracking of one of the two discs using a hand action on the trackpad to control the cursor, while typical stream-bounce displays are passively observed. Third, the data reflect online interpretation of the stream-bounce event, while typical stream-bounce designs provide only a subjective report of the event after its conclusion. Last, our design includes an additional moving stimulus (the cursor).

We conducted three control experiments. The first was completed by the main participant group after they had completed the main experiment, while the other two were completed by a second participant group (the control participant group: 30 undergraduate students unaware of the purposes of the experiment; 19 female, 11 male; age: 19.70 ± 3.65 years). We discuss each control experiment separately, noting that the stimuli and procedures for each were the same as for the
subjective-motion block in the main experiment except where noted.

**Control experiment 1: End selection:**

Rather than making a subjective report (“stream” or “bounce”) after the motion had completed (as is typical in stream-bounce studies), our design implemented continuous active tracking of a target throughout the motion sequence. Although we replicated the typical stream-bounce effect, we further tested our approach to produce the typical effect when active target tracking was replaced by only covert attentional tracking with subjective report after the completion of motion.

Subjective report after the motion sequence was implemented by disabling active tracking during the motion sequence (the cursor was still visible in the start position) and then reenabling tracking once the targets had reached the end point and stopped. Participants were instructed to track the indicated target covertly during motion, wait for the motion sequence to end, and then swipe the cursor to the appropriate disc depending on whether they had perceived a stream or a bounce event.

We computed the percentage of trials yielding bounce responses for each participant, and these were the units for statistical analyses (see Figure 7 for group mean responses for each sound condition). Paired *t* tests revealed that the typical stream-bounce effect occurred with end selection: $M$(No-Snd) = 5%, $SE$(No-Snd) = 1%, $M$(Snd) = 78%, $SE$(Snd) = 5%, $t$(29) = 13.93, *p* < 0.001. We conclude that our design produced the typical stream-bounce effect with active tracking using a moving cursor, with covert tracking and subjective, after-the-event reporting, and with active tracking. We note that active tracking results in fewer bounce responses for sound trials than covert tracking with end selection, but the effect is small (a difference of 17% in bounce responses) in comparison to the stream-bounce effect itself (a difference of 55% in bounce responses between sound and no-sound trials).

**Control experiment 2: Button tracking:**

The second control experiment addressed potential confounds relating to the use of active tracking of the target using a hand action that is congruent with both the cursor motion and the horizontal component of the target’s motion. Here, to control for displacement of the cursor with hand motion, we replaced trackpad control with button press of the left and right arrow keys of a standard computer keyboard. Although the speed of the cursor was fixed (slightly faster than the target speed, to allow for catch-up after delayed reactions), releasing the button stopped the cursor. This
allowed repeated tapping and releasing of the control key to generate varying average velocities.

We first computed the percentage of trials yielding bounce responses for each participant, and these were the units for statistical analyses (see Figure 7 for group mean responses for each sound condition). Paired $t$ tests revealed that the typical stream-bounce effect occurred with button tracking: $M$(No-Snd) = 18%, $SE$(No-Snd) = 5%, $M$(Snd) = 72%, $SE$(Snd) = 6%, $t$(29) = 6.97, $p < 0.001$.

We next examined frame-by-frame tracking velocities. Planned paired $t$ tests comparing the tracking velocity of stream versus bounce responses for each frame of the motion sequence were conducted to determine when the tracking velocities reliably diverged (Figure 8). The first frame to reliably diverge was Frame 59 (i.e., 33 ms before the PoC)—$M$(Str) = 6.2, $SE$(Str) = 0.2, $M$(Bnc) = 5.4, $SE$(Bnc) = 0.3, $t$(26) = 2.35, $p = 0.027$—and after this point, the tracking velocity for stream responses was significantly different from that for bounce responses for all remaining frames until the end of the motion sequence. Our findings that button tracking leads to differential tracking velocities for stream versus bounce responses and that these velocities reliably diverge before the PoC (and well before tracking of objectively moving targets diverge) support our main findings. We note that the divergence is not as early as for congruent hand tracking (the main experiment) but suggest that this is likely due to the much less precise cursor control afforded by buttons in comparison to the trackpad.

**Control experiment 3: Orthogonal tracking and selection:**

The final control experiment had two aims. We further considered the potential confound of direction-congruent hand-motion direction by replacing it with an orthogonal hand-motion direction. We simultaneously considered the relationship between continuous tracking response and subjective, after-the-fact reporting. We controlled for congruent hand-motion direction by rotating the trackpad 90° so that hand-motion direction and cursor-motion direction were orthogonal (the rotation was made so that “away” moved the cursor in the direction of the dominant hand—i.e., for a right-handed user, hand action away from the body moved the cursor to the right). To gauge the relationship between tracking response and subjective percept, participants were instructed at the end of the motion sequence (and tracking response) to indicate their overall impression of the motion sequence (regardless of tracking outcomes) using a standard computer keyboard (S to indicate stream and B to indicate bounce).

We first computed the percentage of trials yielding bounce responses (with tracking) for each participant, and these were the units for statistical analyses (see Figure 7 for group mean responses for each sound condition). Paired $t$ tests revealed that the typical stream-bounce effect occurred with orthogonal tracking: $M$(No-Snd) = 13%, $SE$(No-Snd) = 4%, $M$(Snd) = 67%, $SE$(Snd) = 6%, $t$(29) = 7.70, $p < 0.001$.

We next examined frame-by-frame tracking velocities. We conducted planned paired $t$ tests comparing the tracking velocity of stream versus bounce responses for each frame of the motion sequence to determine when the tracking velocities reliably diverged (Figure 8). For orthogonal tracking, the first frame to diverge reliably was Frame 55 (i.e., 100 ms before the PoC)—$M$(Str) = 5.5, $SE$(Str) = 0.4, $M$(Bnc) = 4.9, $SE$(Bnc) = 0.3, $t$(26) = 2.21, $p = 0.036$—and after this point, the tracking velocity for stream responses was significantly different from that for bounce responses for all remaining frames until the end of the motion sequence.

Our findings that orthogonal hand tracking led to differential tracking velocities for stream versus bounce
responses and that these velocities reliably diverged before the PoC (and well before tracking of objectively moving targets diverged) support our main findings. We note that the divergence does not occur as early as for congruent hand tracking but suggest that the increased complexity of orthogonal motion may simply be making the task more difficult. With further practice, the results for orthogonal motion could approach those for congruent motion.

To examine the relationship between continuous-tracking response and subjective, after-the-fact reporting, we first computed the percentage of trials yielding bounce responses (with button response) for each participant, and these were the units for statistical analyses (see Figure 6 for group mean responses for each sound condition). Paired t tests revealed that the typical stream-bounce effect occurred with button response: $M(\text{No-Snd}) = 15\%$, $SE(\text{No-Snd}) = 4\%$, $M(\text{Snd}) = 62\%$, $SE(\text{Snd}) = 6\%$, $t(29) = 6.99$, $p < 0.001$.

To compare the response methods, we conducted a two-factor ANOVA (response type: tracking or selection; sound: present or absent) and found a significant main effect of sound, $F(1, 29) = 59.2$, $p < 0.001$, $\eta_p^2 = 0.671$, but no significant main effect of response type, $F(1, 29) = 0.3$, $p = 0.564$, $\eta_p^2 = 0.012$, nor a significant interaction, $F(1, 29) = 2.7$, $p = 0.111$, $\eta_p^2 = 0.085$. We additionally conducted a linear-regression analysis on the percentage of bounce responses for all trials (sound and no-sound) for button response versus tracking response across individual participants, and found an $R^2$ of 0.78. We therefore conclude that for our design, active tracking reasonably reflects subjective percept.

These control experiments provide converging evidence that our main findings are due to the stream-bounce effect itself and are not the result of potential confounds relating to our tracking paradigm. Specifically, neither covert nor active tracking of a target disc acts to destroy the stream-bounce effect. Moreover, tracking responses are highly correlated with subjective perceptual report. Finally, our main finding that precoincidence tracking velocity predicts stream-bounce response does not rely on hand action that matches the horizontal component of the target’s motion.

## Discussion

To examine the temporal unfolding of perceptual decisions, we tracked responses to stream-bounce stimuli dynamically during the event rather than collecting a subjective report after the event. Specifically, we compared target-tracking velocities throughout the motion sequence for reported streaming versus bouncing events and made two predictions. First, we considered that tracking a bouncing target requires a reversal of tracking motion following the bounce, whereas tracking a streaming target does not. Thus, we predicted an abrupt and clear divergence in velocity profiles sometime after the point of target coincidence for streaming versus bouncing responses. This reversal in tracking motion would reflect the actioning of a bounce versus a stream response. Second, we contended that if top-down cognitive factors influence perceptual decisions, these factors may manifest as behavioral differences. If top-down factors influence perception early, then we expected a preparatory divergence in tracking velocities for bouncing versus streaming responses prior to the divergence with reversal that indicates the actioning of the stream or bounce response.

We first note that our tracking paradigm provides a valid approach to investigating the stream-bounce effect: The effect itself is robust and does not depend on the response mode. The main experiment (congruent active tracking) and all the control experiments produced the usual effect, that responses are modulated depending on the presence or absence of the sound, and the magnitude of the effect agrees well with the literature (e.g., Sekuler et al., 1997).

Both of our predictions were confirmed. Across both conditions in the main experiment (objective and subjective motion), tracking velocities for bouncing versus streaming responses abruptly and clearly diverged about 250 ms after the targets coincided, with tracking for bounce responses rapidly decelerating to zero and then reversing (negative velocity), whereas tracking for streaming responses remained positive throughout. This divergence reflects the actioning of the stream or bounce response. Importantly, we also observed a preparatory response-related divergence in tracking velocities about 500 ms before the targets coincided and 750 ms before the clear stream-bounce response divergence in the subjective-motion condition only. At this point, tracking for what would become bounce responses slowed significantly compared to tracking for what would become stream responses. In the period between the two points of divergence (i.e., $-500$ to $+250$ ms around the point of target coincidence), the streaming and bouncing tracking velocities followed similar but offset profiles with positive but slowly decelerating velocities. To be clear, no preparatory motion was observed for the objective-motion tracking condition. The two predictions were confirmed both when considering all trials within the experimental block (i.e., visual-only and audiovisual) and when considering only audiovisual trials.

Our main finding is a significant behavioral change preceding a perceptual decision that then predicts that decision; tracking slows down before the point of coincidence on trials that become bounce responses.
compared to those that become stream responses. This pattern of results is compatible with three possible underlying processes. First, it is possible that actions simply drive responses to the stimuli independent of perceptual outcomes. That is, some early (conscious or unconscious) motor decision or action preparation to reverse tracking or not results in a stream or bounce outcome that does not necessarily reflect the subjective percept. Second, it is possible that action drives perception. Perhaps an early (conscious or unconscious) motor decision or action preparation for reversal or not biases the perceptual decision. Finally, high-level cognitive processes possibly influence both action and perception in a top-down manner. We favor this final explanation and suggest that cognitive expectation of an upcoming perceptual outcome biases the interpretation of sensory input to favor the expected percept and simultaneously prepares action for the appropriate response. We discuss each of these possibilities.

Consider the proposal that actions simply drive responses and these responses do not necessarily reflect perceptual decisions. The results of Control Experiment 3 (orthogonal tracking with end selection) suggest that tracking responses reflect subjective percepts. We observed a strong correlation between active-tracking responses and subjective, after-the-event reporting of the same trial as a stream or bounce. Although it is possible that subjective reports were made to be consistent with active-tracking responses, we discount this for two reasons. First, participants were asked to respond subjectively based on their impression of the entire motion sequence and were specifically instructed to ignore their tracking response. Second, bounce responses are strongly associated with the presence of a sound at coincidence. If precoincidence actions drive tracking independently of percept, we would expect tracking responses to be generally independent of the sound at coincidence and not show the sound/no-sound polarization typical of stream-bounce displays.

There is considerable prior research supporting our second proposal, that action drives perception. Gutteling, Kenemans, and Neggers (2011) investigated participants’ sensitivity to a small orientation change of a visual stimulus with either a relevant grasping action or a nonrelevant pointing action. They found that sensitivity increased for the relevant (grasping) action and took this as direct evidence for perceptual enhancement of a specific relevant feature in preparing a motor act. Action-modulated perception has also been shown to facilitate visual search for grasping-relevant features (Bekkering & Neggers, 2002); and Vishton et al. (2007) have proposed that intention to reach for a target influences perception by changing the perceived size of the target. More pertinent to the present study, hand movement biases the perceived direction of ambiguous motion displays. Combining bistable circular apparent motion with rotating hand motion (turning of a knob), Wohlschläger (2000) showed that both congruent action and the planning of congruent action biased perception.

Despite this body of research, we discount the possibility of action driving perception as an explanation of our findings based on three reasons: The results of our control experiments do not support this interpretation; the observed action differences (differential tracking velocities) in these results are extremely subtle compared with typical action manipulations that have been shown to influence perception; and, although relevant actions can modulate perception, there is other evidence suggesting that irrelevant actions such as passive tracking do not. We discuss each of these in detail.

First, Control Experiments 2 and 3 demonstrated that the stream-bounce effect and early preparatory divergence of velocity profiles for streaming versus bouncing responses occur not only for congruent active tracking but also for button-controlled and orthogonal tracking (where hand motion was orthogonal to the cursor motion). Although the action goal for each of these is the same (i.e., tracking control of the cursor), prior evidence for action modulating perception typically involves actions that are themselves in some way congruent with the perceptual decisions to be made. For instance, rotating a knob clockwise or counterclockwise biases rotational motion clockwise or counterclockwise (Wohlschläger, 2000), and grasping modulates the perception of grasping-relevant visual features (Gutteling et al., 2011). Specifically, our findings persist for actions that are incongruent with the perceptual outcomes (orthogonal hand motion) and for actions that are functionally unrelated to the perceptual outcomes (button pressing).

Second, we note that where action modulates perception, the actions themselves (even if only prepared and not actually executed) are typically overt (turning a knob clockwise versus counterclockwise or grasping versus pointing) and long-lasting. The action difference related to differing perceptual decisions here is extremely subtle: less than 10% difference in tracking speed, equivalent to approximately 0.5°/s, lasting for just half a second before the targets coincide. It is not clear that such subtle actions could modulate perception in the same way that more overt actions have been shown to do. Further, where actions modulate perception, the actions are typically cued and then prepared and executed with intention. Again, it is not clear that the tracking action in this study—which is a ballistic, unplanned action with online modifications—would have the same effect on perception.

Finally, we refer to earlier studies that use paradigms more like the present study than those already
mentioned. Using a version of the stream-bounce display where the motion of the targets themselves was controlled using a mouse, Mitsumatsu (2009) found that voluntary action biased perceptual outcomes. In that setup, participants moved a mouse in a streaming action to control the motion of both targets (one target moved with the mouse, the other with corresponding but reversed motion). The streaming hand action significantly reduced bounce responses for audiovisual stimuli. Furthermore, the effect persisted with an orthogonal motion control, where target motion remained left/right but mouse motion was toward/away from the participant. While these results appear to conflict with our claims regarding our controls, when causal control (mouse movement controlled target movement) was compared with noncausal control (mouse movement initiated but did not control target movement) the same streaming mouse action no longer resulted in reduced bouncing for noncausal control. In other words, simply moving with a target but not controlling it had no effect on perceptual outcomes; the action goal of object control was critical. We contend that this is precisely the situation in our setup, where there is hand motion but no control of the targets themselves (only the unrelated cursor).

Ichikawa and Masakura (2006) conducted a similar study using the flash-lag effect (when a stationary flash is paired with a moving target, the flash is perceived in a lagged position) and found similar results regarding the critical nature of causal control. They examined four hand-movement conditions whereby motion of a computer mouse varyingly controlled the motion of the moving target in a flash-lag display: full manual control (hand movement controlled the target), automatic target (no hand movement at all), half automatic (hand movement initiated and briefly controlled the target), and accompanied (noncausal) hand motion. Whereas the full-manual and half-automatic control conditions reduced the flash-lag effect, neither the automatic-motion nor noncausal-hand-motion conditions did. Following up with a second study (Ichikawa & Masakura, 2010), they found that while congruent mouse motion with control reduced the effect, incongruent motion did not, nor did control using a key press or trackball. They concluded that manual operation to control the stimuli facilitates visual perception and reduces the effect, but incongruent and noncausal actions do not affect perception. Again, because our setup most closely resembled the accompanied (noncausal) hand motion, action is unlikely to be driving perception.

So although previous work demonstrates that actions can influence perception, these actions are typically obvious whereas ours are subtle, causal whereas ours are noncausal, and they generally require congruent motion whereas our effect persists for two types of incongruent motion. We propose that our findings are not due to action influencing perception but are the result of a simultaneous top-down influence on both action and perception.

Parise and Ernst (2017) investigated the roles of noise and previous response in the perception of stream-bounce stimuli and found a strong previous-trial effect. In addition to the usual finding that sound strongly modulates responses, they found that the previous response also modulated responses, and to a similar degree to sounds (the previous-trial modulation was a positive effect reflecting perceptual stabilization). Our findings agree with these. Their main analysis, however, consisted of introducing visual noise into the stream-bounce stimulus and then determining (by reverse correlation) whether there was any signal in the noise that correlated with perceptual outcomes. Certain visual information prior to target coincidence was associated with increased bounce responses. Importantly, although both sound and previous trial modulated responses, neither was associated with the early visual information that increased the likelihood of a bounce response. The authors suggested that these findings argued against both sound and perceptual memory influencing early visual processing. While it is possible that the precoincidence velocity effect we observed may reflect a similar influence of noise in the visual system biasing outcomes, we suggest that there are two important differences between our study and Parise and Ernst’s that need to be considered.

First, signal to noise. Parise and Ernst used visual stream-bounce stimuli that were embedded in dynamic visual noise with a relatively low signal-to-noise ratio (the luminance of each noise sample varied between 14.6 and 48.3 cd/m², while the targets were 50 cd/m²). The stimuli in our study, on the other hand, were of high contrast (black on a midgray background). While internal (neural) noise would be present in both stimuli, the signal-to-noise ratio in our study is not only much greater but likely orders of magnitude greater than in the Parise and Ernst study. That noise can bias the interpretation of a weak stimulus does not necessarily imply that it can measurably bias the interpretation of a strong stimulus.

Second, while Parise and Ernst found that precoincidence noise could bias perceptual disambiguation, it is not obvious that noise could therefore simultaneously bias action. Certainly, visual processing and visual perception influence action, but this does not imply that noise in early processing, acting from the bottom up, influences action before the perceptual outcome. Considering these two issues and the other evidence already discussed, we maintain that the best interpretation of our findings is that cognitive factors are simultaneously biasing both action and perception.
The idea that action and perception may be prepared early is not new. A bilateral slowly increasing negativity (termed the “readiness potential”) was shown to precede voluntary action as early as 1965 by Kornhuber and Deecke. Libet (1985) famously extended this work to show that the readiness potential precedes not only a voluntary endogenous motor act but even the intention to act by hundreds of milliseconds. This suggests that the initiation of a spontaneous action begins unconsciously before subjective awareness, and the conscious decision to act follows. More recently, using fMRI, Soon, Brass, Heinze, and Haynes (2008) demonstrated that the content of motor decisions was encoded in prefrontal and parietal brain activity up to 10 s before conscious awareness, and the timing of those decisions was predicted 5 s early in the supplementary motor area and pre-SMA.

Early brain activity has also been shown to predict perceptual decisions. Hesselmann, Kell, Eger, and Kleinschmidt (2008) found that prestimulus activity in the fusiform face area predicted perceptual decisions of an intermittently presented Rubin vase, with increased activity associated with a face interpretation. This was then extended (Hesselmann, Kell, and Kleinschmidt, 2008) to an ambiguous motion stimulus, and prestimulus activity in hMT+ predicted whether close-to-threshold coherent motion was perceived as coherent. In a visual-discrimination task, Bode et al. (2012) presented participants with varying degraded images of pianos or chairs mixed with occasional pure-noise images. Although poststimulus electroencephalographic activity predicted the discrimination of real stimuli (pianos or chairs), activity during the 100 ms before stimulus presentation predicted piano or chair responses to pure noise.

Importantly, two imaging studies using stream-bounce displays have found that prestimulus activity can predict perceptual outcomes. Recently, Zhao et al. (2017) considered response-associated (stream or bounce) electroencephalographic activity to just visual-only stream-bounce stimuli. They found that early visual event-related potentials were no different for streaming versus bouncing responses, and that postcoincidence activity over parietal scalp dissociated perceptual responses. Their main finding, however, was that frontocentral activity 200 ms prior to coincidence (P2) predicted perceptual outcomes. They proposed two possible explanations for this activity leading to responses: It could be associated with a cognitive matching system that compares sensory input with expectations, or—noting that a similar positivity (P2) has been shown to be sensitive to depth perception—the larger P2 for streaming responses could reflect an early decision or assumption that the targets in a specific trial are in different depth planes. Either way, some expectation or prediction regarding the ambiguous stimuli is acting as an early influence on the perceptual decision.

Hipp, Engel, and Siegel (2011) considered just audiovisual stream-bounce stimuli and investigated electroencephalograph spectral synchronization across brain regions. Beta-band synchronization across frontal and parietal regions during the 125 ms before coincidence and gamma-band synchronization across sensorimotor, premotor, and temporal regions starting before and peaking around target coincidence predicted subjects’ percepts.

By presenting participants with intermixed trials of both visual-only and audiovisual stimuli, our findings extend each of these stream-bounce imaging studies. While each of the imaging studies demonstrates the early influence of top-down factors, the stimulus is unchanging. In the present study, visual-only trials were almost exclusively associated with stream responses. Two things were required for a bounce response: an early, preparatory action and a sound. Our findings therefore demonstrate the combined influence of top-down factors and bottom-up sensory input on reaching a perceptual decision.

To summarize: Using a valid stream-bounce design we found behavioral evidence for top-down cognitive factors simultaneously influencing both a perceptual decision regarding an event and a related action a full half second before the critical moment that defined the event and three-quarters of a second before initiation of the action. From this, we draw three conclusions. First, top-down cognitive factors exert an influence on perceptual decisions that is early and can precede the stimulus to be perceived. Second, top-down cognitive factors that exert a prestimulus influence on perceptual decisions exert a similar influence on actions that will ultimately depend on those decisions. Third, in some circumstances, action may provide a measure of perceptual decision making that is independent of subjective report and reflects both the development of the decision over time and the decision itself.

These findings and conclusions are compatible with a large body of literature on early top-down influences on perception and action. We further suggest that they are compatible with two theories regarding perception and action: predictive coding and common coding. Predictive coding (Friston, 2003; Rao & Ballard, 1999; Summerfield et al., 2006) is a theory of sensory processing positing that the brain generates hypotheses or internal models regarding the sensory environment and uses these to actively predict the immediate sensory input. Rather than intensively process the detailed incoming sensory input, low-level processing involves comparing the sensory input with the top-down predictions and transmitting only prediction errors, which then update the internal models. The common-coding framework (Hommel, Müsseler, Aschersleben,
& Prinz, 2001; Prinz, 1997) explains action/perception interactions. It assumes that planned actions and perceived events share a common representational medium, and predicts that action and perception codes induce each other due to their overlap.

We suggest three directions for future research based on our findings. First, a broad range of top-down cognitive factors may influence perceptual decision making, including prior knowledge (e.g., Kersten et al., 2004), expectations (e.g., Gau & Noppeney, 2016), preferences (e.g., Dunning & Balcetis, 2013), and intentions (e.g., Pitts, Gavin, & Nerger, 2008; Suzuki & Peterson, 2000). Continuous-tracking approaches that manipulate specific top-down factors to tease apart and isolate their relative influence would be useful. Second, while previous stream-bounce imaging studies have found early effects that predict perceptual outcomes (Hipp et al., 2011; Zhao et al., 2017), each of those studies considered perceptual bistability in the context of a single stimulus type only (either all visual or all audiovisual). We suggest that extending imaging studies to mixed presentation of visual-only and audiovisual stimuli may provide insight into the influence of top-down factors on multisensory integration beyond perceptual bistability. Third, we suggest that our active-tracking paradigm could provide a precise way to approach investigations into the temporal aspects of other perceptual phenomena, link perception or cognition with behavior, or simply replace or augment subjective report.

Keywords: cognitive inference, multisensory integration, predictive coding, common coding

Acknowledgments

Commercial relationships: none.
Corresponding author: Mick Zeljko.
Email: m.zeljko@uq.edu.au.
Address: School of Psychology, The University of Queensland, Brisbane, Australia.

Footnotes

1 As it is the relative rather than absolute characteristics of tracking that are of interest in this analysis, the units are essentially arbitrary, and so for the sake of simplicity they have been expressed as pixels (rather than degrees) for distance and frames (rather than seconds) for time.

2 We selected 250-ms epochs based on the characteristics of the tracking-velocity profiles. First, we note that the profiles contain three natural epochs: the period spanning 0–500 ms (Frames 1–31), wherein tracking velocity is identical for streaming and bouncing responses in both the objective- and subjective-motion blocks; the period spanning 500–1,250 ms (Frames 31–76), wherein tracking velocities have diverged for streaming versus bouncing responses in the subjective-motion block but not the objective-motion block, and a definitive response divergence has not yet occurred; and the period spanning 1,250–2,000 ms (Frames 76–121), wherein there is a clear response divergence between streaming and bouncing responses in both the objective- and subjective-motion blocks. We next note that these epochs are of different time lengths (500, 750, and 750 ms), and the second epoch spans the critical PoC. The largest duration that allows for equal-length epochs that do not overlap natural epochs is 250 ms; this also avoids an epoch that spans the PoC.

References


