Dissociating low-level visual features from high-level event structure in action segmentation

Zekun Sun & Samuel D. McDougle
Yale University

Version: 5/8/25
Word Count: 5510

Demo page:
https://zk.actlabresearch.org/segmentation
OSF repository:
https://osf.io/j85fa/?view_only=5e8b591d440f49c4ba56150befe9e680

Zekun Sun (*zekun.sun@yale.edu*)
Samuel D. McDougle (*samuel.mcdougle@yale.edu*)
Department of Psychology
Yale University
100 College St.
New Haven, CT 06510

Abstract

Event segmentation is a fundamental component of human perception and cognition. The growing field of event cognition studies how people decide where events occur in incoming sensory data, how "event boundaries" alter decision-making and memory processes, how events reveal themselves in neural activity, and how events may be represented within perception itself. That last point is critical — the representation of events in the first place is filtered through perception. But there is a key open question in the field: Is the perceptual representation of events a simple reflection of the fact that event boundaries are accompanied by large changes in low-level visual inputs (e.g., a sudden cut in a movie scene)? Or, do our higher-level internal models of events (e.g., "step one" versus "step two" of a tennis serve) shape how events are perceived? Here, across seven preregistered experiments, we attempt to dissociate the roles of lower-level visual features and higher-level semantic structures in perception of event boundaries. First, participants produced boundary labels by segmenting brief physical actions (e.g., kicking a ball). Then, separate groups of observers were asked to visually detect subtle disruptions in the actions at boundary versus non-boundary timepoints. The results consistently showed an interfering effect of event boundaries on the detection of disruptions. Critically, boundary effects were strongest when stimuli were presented in recognizable forms versus distorted forms that only preserved lower-level features. Thus, automatic and rapid perceptual segmentation of observed actions may be influenced by both sensory cues and our internal models of the world.
Keywords: event segmentation; observed actions; spatiotemporal dynamics; motion

Public significance statements

Event segmentation refers to the mind's ability to represent continuous sensory input as discrete units unfolding over time. This process might occur automatically within perception before we deliberately think of where to set event boundaries. For example, we might naturally perceive a single golf swing as consisting of a backswing followed by a downswing — two discrete events. How does the mind identify event boundaries? Does it rely on detecting salient features at the transition of events — like how we visually segment objects in continuous space? Or might higher-level information shape event perception? This study explored these questions in a case study of single actions. Our results suggest that the perception of event boundaries is shaped not only by low-level visual cues, but also by higher-level knowledge about the structure of actions.

1. Introduction

The human mind tends to represent continuous experiences as discrete events, imposing "event boundaries" on incoming streams of sensory data. This phenomenon, known as event segmentation, is not just a reflection of explicit knowledge (e.g., "Sally was sitting, now she is standing") – event boundaries appear to have special status within perception itself (Zacks et al., 2007).

Ample evidence has suggested a perceptual root of event segmentation in the mind: Event segmentation first happens in perceptual systems, where the mind appears to spontaneously impose "breakpoints" that segment continuous perceptual experience into a sequence of events with a defined start and end (Kurby & Zacks, 2008; Zacks, 2020). Neural data also show robust activity at event boundaries across a range of cortical sensory and association regions when observers passively view stimuli with clear event structure (Antony et al., 2021; Pomp et al., 2024; Speer et al., 2003; Zacks et al., 2001, 2006). A number of psychophysical studies have provided more direct evidence that event segmentation is baked into perception itself. For example, the representation of the transition between events interferes with the perceptual detection of subtle disruptions, resulting in a lower detection accuracy in detecting disruptions at boundary versus non-boundary time-points within events, in both visual perception (Huff et al., 2012; Y. Ji & Papafragou, 2022; Yates et al., 2024) and auditory perception (Repp, 1992, 1998; See also Ongchoco et al., 2023a and Goh et al., in press for other types of perceptual effects driven by event boundaries).

What is the nature of such perceptual effects at event boundaries? Here, we asked whether such effects only reflect perceptual sensitivity to changes of stimulus features at the transition of events, or if they are also driven by internal representations of event structures. To our knowledge this question has yet to be directly addressed, as previous studies of event boundaries have been deliberately designed around salient changes of features at event boundaries, such as walking through a doorway (Ongchoco et al., 2023b; Radvansky, 2012; Radvansky & Copeland, 2006), large shifts in the appearance and location of objects and agents (DuBrow & Davachi, 2013; Huff et al., 2012; Pomp et al., 2024; Tauzin, 2015), abrupt changes in scenes (Baker & Levin, 2015; Cutting, 2014; Lee & Chen, 2022; Yates et al., 2022), robust motion cues (Hard et al., 2006; Hemeren & Thill, 2011; Newtson et al., 1977; Ongchoco & Scholl, 2019; Yates et al., 2024; Zacks et al., 2009), and sudden disruptions of visual statistics (Avrahami & Kareev, 1994; Baldwin et al., 2008; Buchsbaum et al., 2015; Ezzyat & Clements, 2024; Schapiro et al., 2013). When there is a dramatic sensory

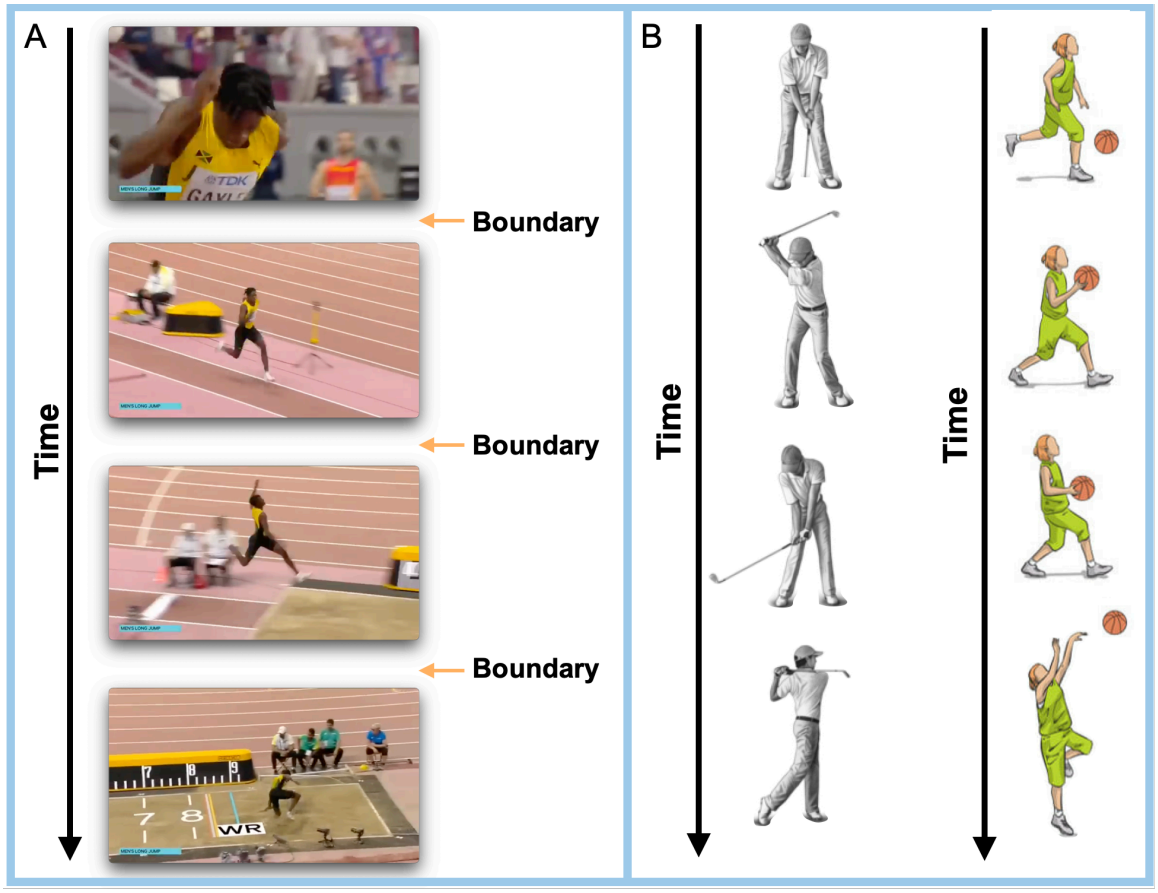

Figure 1: (A) Mental representation of event boundaries emerges from ongoing perceptual processing of continuous sensory data. As shown here, a long jump video is segmented at the shot cut and at the transition between run to jump and jump to stand — timepoints that coincide with abrupt changes in visual features. Is perception of boundary sufficiently driven by visual statistics? Or does it also rely on our internal model of the world? (B) Here, we explore these possibilities for the case of *single actions*, such as a golf swing and a basketball shot. We ask whether and how the discrete 'steps' within a single action are represented in visual perception.

change at the transition of two events, the perception of boundaries could primarily reflect the resulting salient changes in lower-level visual statistics. That is, a cut between movie scenes, a person moving from the kitchen to the dining room, or a ball suddenly accelerating are all accompanied by significant changes in lower-level visual features.

Prior studies that used more abstract, simple stimuli in segmentation tasks (e.g., the videos of moving geometric shapes) have revealed an even stronger relation between boundary judgments and low-level features (such as movement and motion features and dynamics) compared to semantically meaningful videos, accompanied by increased activity in motion-related brain areas at event boundaries (Hard et al., 2006; Pomp et al., 2024; Zacks et al., 2006, 2009). With motion cues as one key factor driving event perception (Speer et al., 2003), it remains unknown whether high-level event structure is also represented in perception and thus contributes to perceptual effects at event boundaries.

Higher-level event structure, e.g., "step one" versus "step two" of a tennis serve, could in theory be represented in perception, shaping continuous perceptual experience into distinct events. A growing body of evidence has suggested the existence of high-level structures in visual perception that are not given directly by stimuli, such as the internal structure of objects (Ayzenberg & Lourenco, 2022; Feldman & Singh, 2006; Sun & Firestone, 2021), thematic roles of agents (Hafri et al., 2018; Papeo et al., 2024), event categories (H. Ji & Scholl, 2024; Strickland & Scholl, 2015), and abstract relations between objects (Hafri et al., 2024; Lovett & Franconeri, 2017; Sun et al., 2025). Yet, to our knowledge, the evidence for high-level event structure in perception of naturalistic, continuous actions is limited.

The present study: Event boundary in single actions

In this work, we ask to what extent the low-level visual features versus high-level internal structure of events contribute to visual segmentation in a case study of observed actions. To foreshadow the key result, perceptual boundary effects were significantly stronger when the high-level structure of observed actions remained perceivable, pointing to a role for internal models of temporal structure in the perceptual representation of events.

Transparency and openness

This research received approval from Yale University's local ethics board. For all experiments, we pre-registered the sample size, experimental design, and the statistical analyses. Demonstrations of all experiments and full set of dynamic stimuli can be viewed at:

https://zk.actlabresearch.org/segmentation.

The data, experiment code, stimuli, and experiment pre-registrations for all studies are available at:

https://osf.io/j85fa/?view_only=5e8b591d440f49c4ba56150befe9e680.

2. Experiment 1: Explicit segmentation

Experiment 1 aimed to identify the most characteristic boundary frame of the individual actions. We asked observers to deliberately segment a variety of action videos into two semantically distinct units. The simple MoCap animations used here appear as short, fluent and natural actions, with no obvious changes of scenes, agents, or objects. We expected participants' boundary judgments to reflect how they explicitly represent the discrete steps of observed actions in a top-down manner (e.g., "step one" and "step two" of a tennis serve).

2.1. Method

2.1.1. Participants

As stated in our pre-registration, we recruited 100 participants from the online platform Prolific (https://www.prolific.co/). Participants were pre-screened for a minimum approval rate of 99%, at least 50 prior submissions, normal or corrected-to-normal vision, fluency in English, and U.S. residence.

2.1.2. Stimuli

We compiled 20 animations depicting short, natural actions, spanning sports, simple exercises, and everyday tasks, from the CMU *MoCap* database(http://mocap.cs.cmu.edu/). These videos involve an unidentified and skeletonized human figure performing a single, well-defined action (e.g., making a golf swing, picking up an object, taking a step aside) on a black background (Figure 2). Such simple, short actions have a salient basic structure, and are thus often referred to as "bounded events," which reflect events that lead to a salient start and endpoint that is naturally achieved unless there is a surprising interruption (Comrie, 1976; Y. Ji & Papafragou, 2022; Mittwoch, 2013). The approximate duration of the 20 actions ranged from 1.0 s to 3.9 s (Mean = 2.2 s, SD = 0.7 s).

The action videos (200 × 200 pixels) were displayed in participants' web browser. The workspace covered 500 × 500 pixels with a black background. Because of the nature of online studies, we could not know the exact viewing distance, screen size, and luminance (etc.) of these stimuli as they appeared to participants. However, any distortions introduced by a given participant's viewing distance or particular

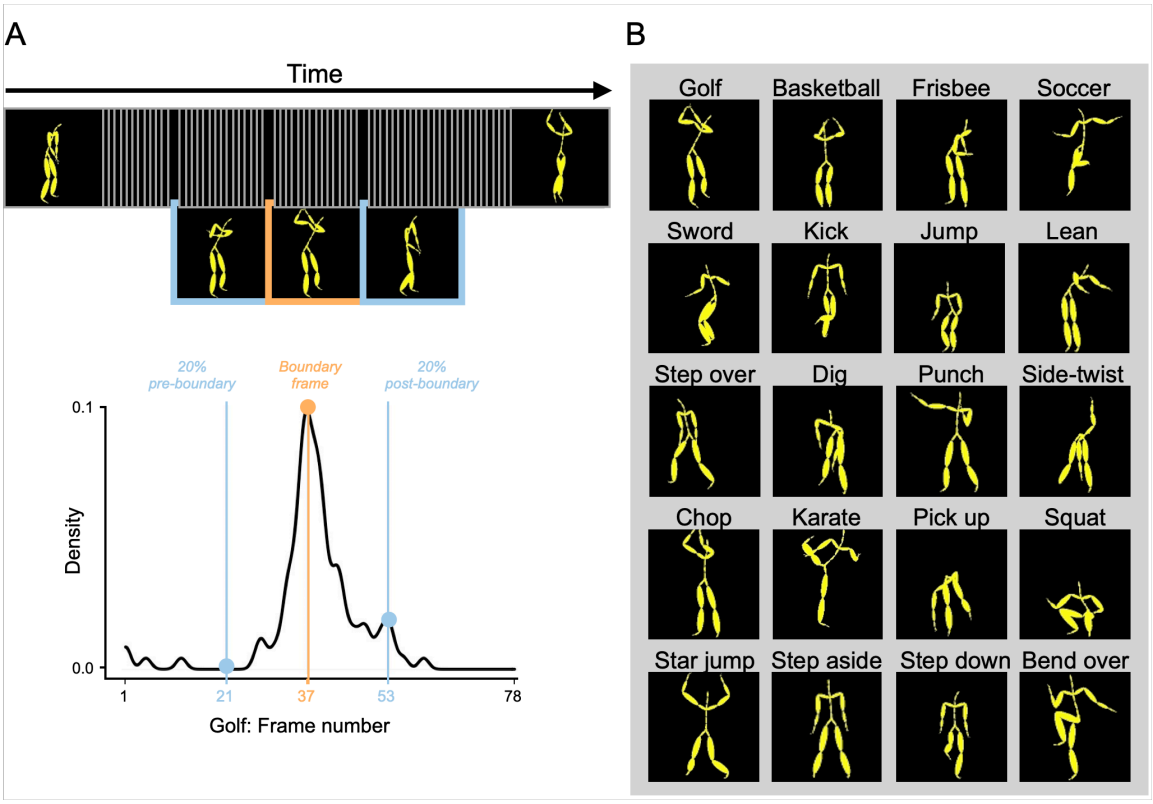

Figure 2: (A) The method used to define the most characteristic event boundary for an action. Taking the golf swing video as an example, all event boundary responses from participants were combined and smoothed with a Gaussian kernel. The frame closest to the stimulus' global peak of the event boundary choices was identified as the boundary frame, and 20% frames of the video clip before the boundary as the pre-boundary frame and 20% after the boundary as the post-boundary frame. (B) The 20 actions used in this study. Each image represents the boundary frame of each action video, as determined by our analysis method in Experiment 1.

monitor settings would have been equated across all stimuli and conditions. No time pressure was applied to participants' responses.

2.1.3. Procedure

Each trial started with a "ready" cue appearing at the center of the workspace, which reminded participants to observe the full video of an action. After viewing the full video, they were given a slider they could toggle with their mouse, which allowed them to iterate through all the frames of the previously viewed action. Their job was to move the slider in order to find the most appropriate frame that divides each action into two units, such that each unit is still "meaningful." The width of the slider was determined by the number of frames of the current action video, such that the amount of mouse movement required for seeing each frame was equal across all actions. The starting position of the slider was randomized on each trial. Each participant made these boundary judgments for each of the total 20 unique actions. The serial order of actions was randomized across participants.

2.2. Results

Two participants were excluded for failing to submit a complete data set, leaving 98 participants with analyzable data. For each action, participants' choices of the boundary were combined and smoothed with a Gaussian kernel (bandwidth = 1 frame), which gave us the density of choices for each frame of the action video. The boundary of an action was defined as the frame corresponding to the highest peak of the fitted density function. Using this procedure, we obtained a single event boundary frame for all 20 actions (See an example in Figure 2A). Observers showed high agreement in selecting boundaries: On average, 40.3% observers chose either the group-level peak frame or its neighboring frames as an action's event boundary.

Boundary selections were not limited to the videos' middle frames: The earliest boundary was 29.4% from the beginning across all videos, and the latest was 61.3%. Indeed, as predicted given our instructions, observers typically selected boundary frames that best described how to perform an action in discrete steps, e.g., kicking a soccer ball is divided into planting the supporting foot and then swinging the kicking foot, performing a simple jump requires bending the knees and then extending upward, etc. (Figure 2B depicts each boundary frame.)

3. Experiment 2: Temporal change detection

Experiment 2 explored whether the action boundary was also represented in visual perception, rather than just reflecting an explicit decision about where the boundary

belonged. Previous empirical evidence has shown that perceptual processing of event transitions (i.e., boundaries) transiently impairs one's ability to detect subtle changes in continuous perceptual input (Huff et al., 2012; Y. Ji & Papafragou, 2022; Repp, 1992, 1998; Reynolds et al., 2007; Yates et al., 2024; Zacks et al., 2007)). Here, we instructed observers to detect a transient slowdown in action videos, and tested whether perceptual sensitivity differs between boundary frames versus non-boundary frames.

3.1. Method

3.1.1. Participants

In line with our pre-registration, 20 participants were recruited through Prolific. A smaller pilot suggested that this sample would have power above 95% to reveal boundary effects.

3.1.2. Stimuli and procedure

Participants were instructed to detect transient slowdowns in the animated actions used in Experiment 1. On each trial, observers first watched the full video of an action, followed by a mask image (a yellow random noise pattern) for 900 ms, and then the video was played a second time. The second play was either identical to the first play or contained a transient slowdown (a 60- or 90-ms pause on two-thirds of trials). Observers pressed either F or J on their keyboard to indicate whether the second play did or did not contain a brief pause compared to the first play, with no time pressure to respond (Figure 3A). Critically, the pause occurred at one of three specific frames determined by the results of Experiment 1: the pre-boundary frame (20% of frames earlier than the boundary), the boundary frame, or the post-boundary frame (20% of frames later than the boundary). The three frame conditions were crossed with the 3 pause time conditions, yielding 3 (pre-, on-, post-boundary) $\times$ 3 (0, 60, 90 ms), or 9 unique trial types. All 20 actions appeared as each of these trial types, for $20 \times 9 = 180$ total trials. Trial order was randomized across participants. Participants were also given 3 practice trials at the onset of the task (using an action video that did not appear elsewhere in the experiment).

3.2. Results

Four participants were excluded for low accuracy ($< 60\%$; not significantly above chance according to a binominial test with $\alpha = .05$) and 1 for failing to provide complete data, leaving 15 participants with analyzable data. Across all participants, the mean accuracy was 75.3%. We only analyzed the trials with a pause (60 ms and 90 ms).

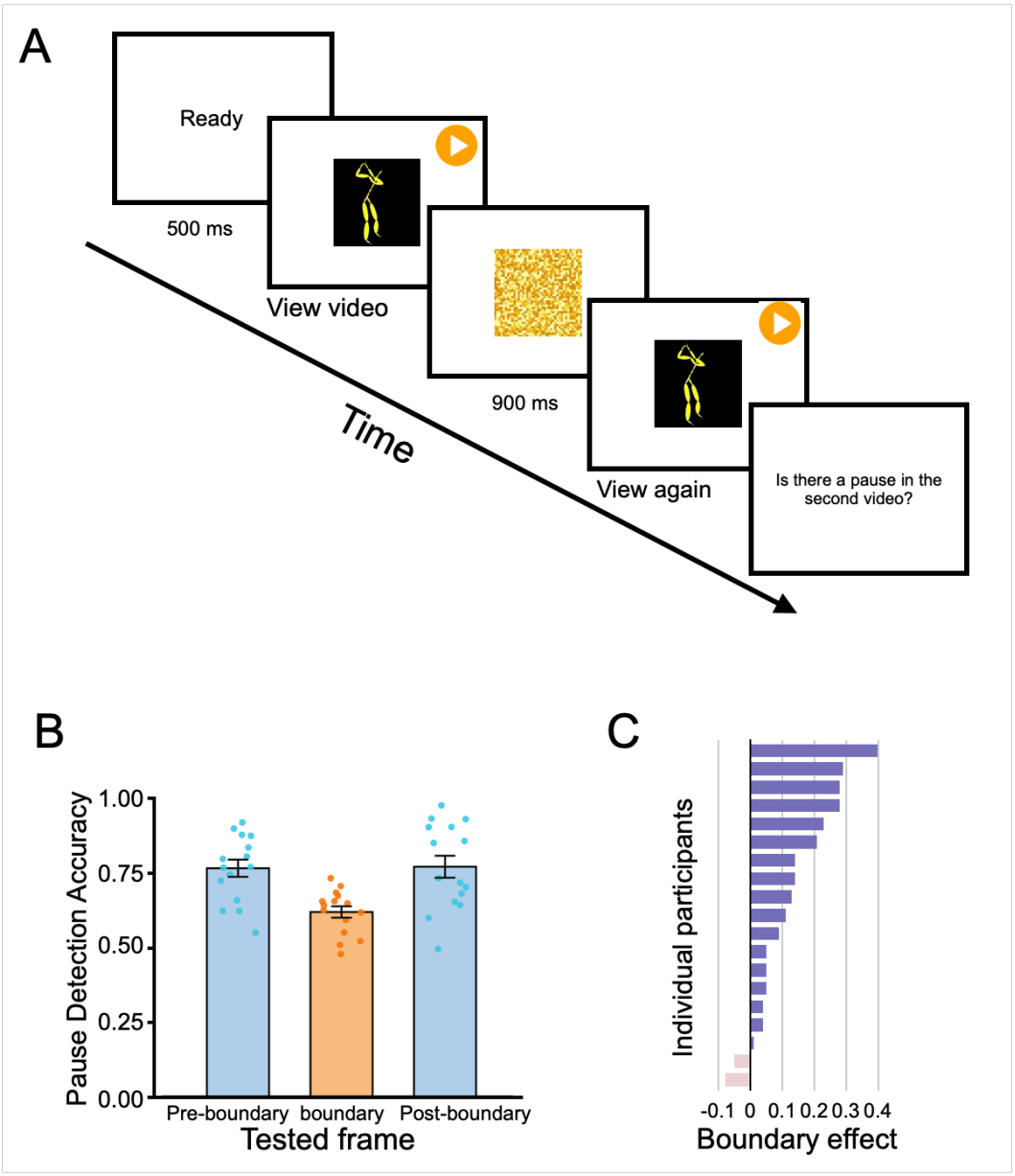

Figure 3: (A) In Experiment 2, observers watched each action video twice. During the second viewing, a brief pause could happen at one of the three tested frames (pre-boundary, boundary, or post-boundary frame). Observers judged whether there was a pause or not in the second play. (B) Compared to the pre-boundary and post-boundary frames, observers were less likely to identify that the pause occurred at the boundaries. (C) A strong majority of participants were less sensitive to the disruptions at the boundary of actions. Error bars = 1 s.e.m.

As shown in Figure 3B, observers were less accurate in detecting pauses at boundary frames relative to non-boundary frames (averaging pre- and post-boundary frames), $t(14) = 4.52$, $p = 0.00049$, Cohen's $d = 1.17$, 95% $CI_{effect} = 0.15[0.084, 0.21]$. 13 out of 15 participants (87%) showed lower visual sensitivity to subtle disruptions at action boundaries. These results suggest that the transition of events within each action was spontaneously represented by the visual system, and, given the simple nature of the stimuli (e.g., a single continuous scene with no objects, perspective changes, or dramatic transitions), implies that these effects may not be fully driven only by dramatic changes in visual features.

4. Experiment 3: Spatial change detection

The previous experiment addressed visual segmentation using dynamic stimuli (i.e., brief videos of actions). However, evidence exists showing that the mind can also form rich event representations from a series of discrete images, where event segmentation cannot rely on motion or dynamic signals (Baldwin et al., 2008; Cohn, Holcomb, et al., 2012; Ezzyat & Clements, 2024; Zheng et al., 2020). For example, people segment narratives while reading the static images of a comic book in a certain order (Cohn, Paczynski, et al., 2012). Here, we tested if perceptual boundary effects similar to those reported in Experiment 2 can emerge from such a scenario, where no motion signals can be used. In this experiment, we asked participants to detect subtle changes of spatial features instead of temporal disruptions in the previous experiment. We predicted that the detection of spatial change would also be weakened at boundary frames if the transition between action steps is represented in perception.

4.1. Method

4.1.1. Participants

As stated in our pre-registration, 20 participants were recruited through Prolific.

4.1.2. Stimuli and procedure

This experiment asked participants to view a sequence of discrete images taken from the original action videos (maintaining the temporal order) and then judge whether the last image shown changed from the previous image. The images were selected from the frames of the action videos used in Experiments 1-2, selecting one out of every three frames to create the sequence. We sequentially presented one frame from every three frames before the designated "key" frame, which was either the boundary, pre-boundary, or post-boundary frame obtained in Experiment 1 (e.g.,

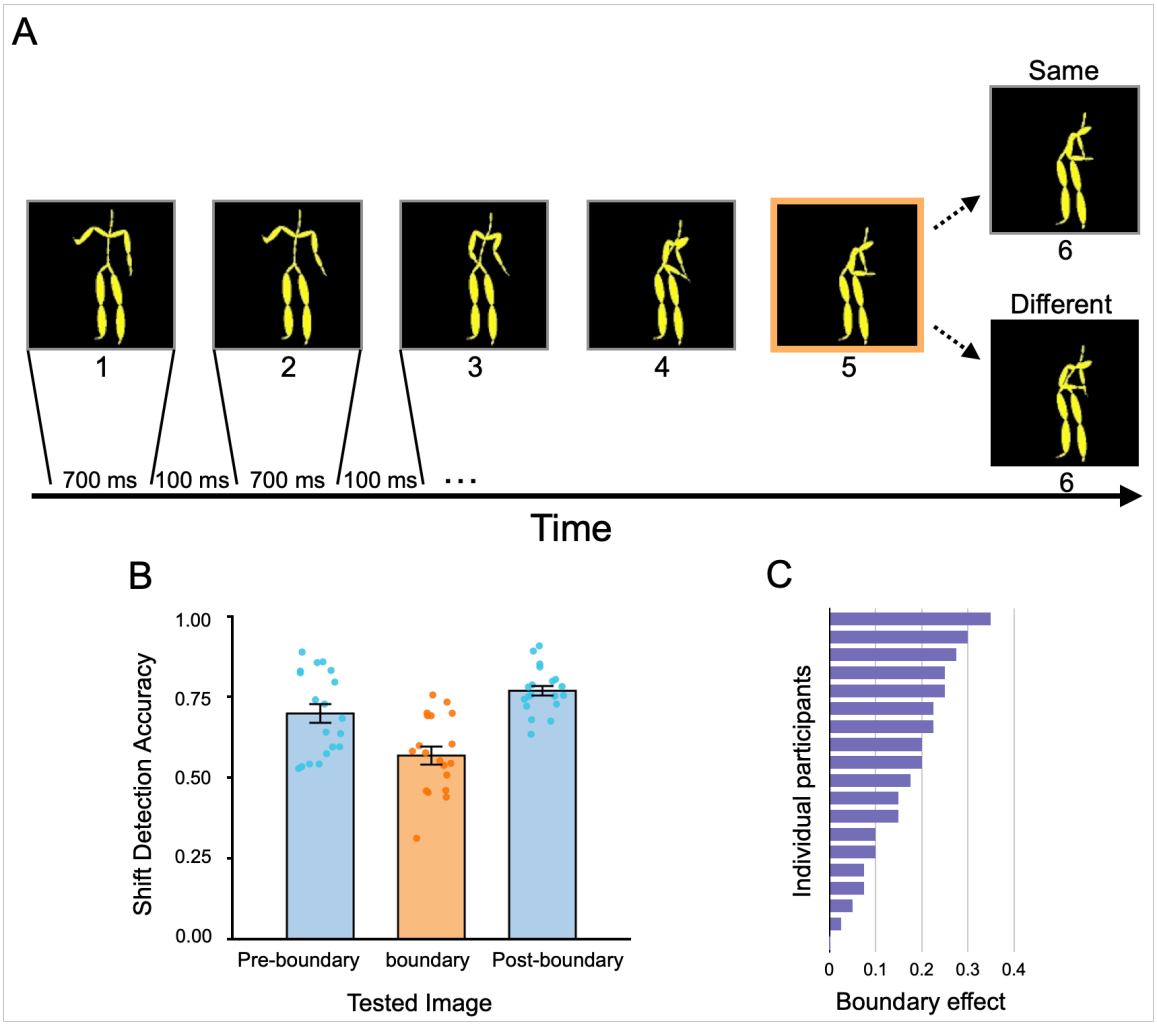

Figure 4: (A) In Experiment 3, participants observed a sequence of images (i.e., individual frames of each action video) turning on and off, with 700 ms duration and a 100 ms ISI. Without knowing how many images would show up, participants were asked to judge if the last image shown changed from the previous image or not. Critically, the second-to-last image on each trial (here, the fifth frame) could be the pre-boundary, boundary, or post-boundary frame of the action. (B) When a change occurred at the boundary frame, participants were less likely to identify it, similar to Experiment 2. (C) All participants, except one, were less sensitive to changes of boundary frames compared to non-boundary frames. Error bars = 1 s.e.m.

in Figure 4A, the orange bordered frame is the boundary frame of the action "tossing a frisbee"). Following the key frame was a testing frame, which was either the same as the key frame, or 2 frames forward in time from the key frame. Thus, from the viewpoint of the observer, on each trial, a series of images was sequentially displayed (each for 700 ms, with a 100 ms ISI), and they simply needed to determine whether the last frame changed from its previous frame or not (Figure 4A). Since observers were not informed about how many images would be shown on each trial, they had to monitor every change of the images.

On each trial, participants pressed a key to observe the sequence of images depicting each action. After the last image was displayed, participants pressed either F or J on their keyboard to indicate whether the last image had shifted at all relative to the previous one. Two factors were fully crossed within-subject: 3 (pre-, on-, post-boundary) $\times$ 2 (shift, no shift) $\times$ 20 (actions) + 4 practice trials = 124 trials. Trial order was randomized across participants (except for practice trials).

4.2. Results

One participant was excluded for low accuracy ($< 60\%$, not significantly above chance), leaving 19 participants with analyzable data, with a mean accuracy of 73.7%. Across 20 actions, observers' detection accuracy at pre-boundary, boundary, and post-boundary frames was 70.8%, 57.6%, and 77.9%, respectively. Disparate performance between boundary and non-boundary frames was confirmed by a paired t-test, $t(18) = 7.41$, $p = 7.14 \times 10^{-7}$, $d = 1.70$, 95% $CI_{effect} = 0.17[0.12, 0.21]$ (Figure 4B). All participants, except one, showed lower sensitivity to image changes at action boundaries compared to non-boundary frames (Figure 4C). The results suggest that the perception of action boundaries interrupted the processing of subtle image differences, even when viewing static stimuli. Unlike Experiment 2, observers made their responses here in the absence of any image motion signals, and were never given a chance to preview the complete action videos. Yet, when the feature change happened around the boundary frame, they were less likely to detect it.

5. Experiment 4a: Low-level dynamic features

Experiments 2-3 revealed a perceptual effect wherein participants were less likely to detect subtle changes at action boundaries compared to non-boundaries. However, are such effects merely driven by low-level difference between boundary and non-boundary frames? For example, if frame-to-frame changes are more subtle around action boundaries, transient slowdowns at boundary frames could be less noticeable than those at non-boundary frames. In this experiment, we invented a new method

— a "spiralized" stimulus — whereby pixel-level dynamics were sufficiently preserved while high-level information was removed from the stimuli, allowing us to ask whether boundary effects were fully explained by lower-level information or not.

5.1. Method

5.1.1. Participants

In line with our pre-registration, 35 participants were recruited through Prolific. A power analysis on the main results of a smaller pilot (i.e., the pre-registered paired t-tests on the boundary effect) suggested that this sample would have power above 95% to reveal the main difference between normal and spiralized videos in terms of boundary effect (considering accidental loss of data). For Experiments 4a–b and 5a that used the same number of conditions and trials, we pre-registered the same sample size.

5.1.2. Stimuli and procedure

We used the same videos but distorted all the frames of the actions with a spiral-shaped filter, generating a new set of videos that preserved lower-level motion signals in the original videos while removing higher-level semantic information. The spiralized videos were created by taking each frame of the given standard action video clip and running it through a twirl filter with 600° using Adobe Photoshop (Figure 5A).

Indeed, the distorted videos maintained a nearly identical frame-to-frame profile in terms of the pixel changes across pairs of frames as the original actions. For example, in the golf swing stimulus, the obvious change in pixel movements between the backswing and the downswing is shown as the transition from clear outward to inward motion in its spiralized version, and the deceleration and acceleration during the swing are similarly perceivable in the spiral pattern. In fact, pixel changes across consecutive frames were highly correlated between original and spiralized videos ($r = 0.97$ across 20 actions, Figure 5B), giving rise to highly similar impressions of motion and dynamics. In some cases, the motion signal was even stronger in the spiralized videos due to the stretching of the pixels (on average, the between-frame shift is 3824 pixels in normal videos versus 4350 pixels in spiralized videos). Readers can experience the similar dynamics in normal and spiralized videos at: https://zk.actlabresearch.org/segmentation/skeleton.html.

Observers attempted to detect transient pauses in the second play of each video. The experimental design was the same as Experiment 2, except that: 1) two types of videos were used – normal videos versus spiralized videos, and 2) for each participant, 10 randomly-selected actions used 60 ms for the pause trials and the other 10 used 90 ms. The Video Type condition was crossed with the other conditions as in

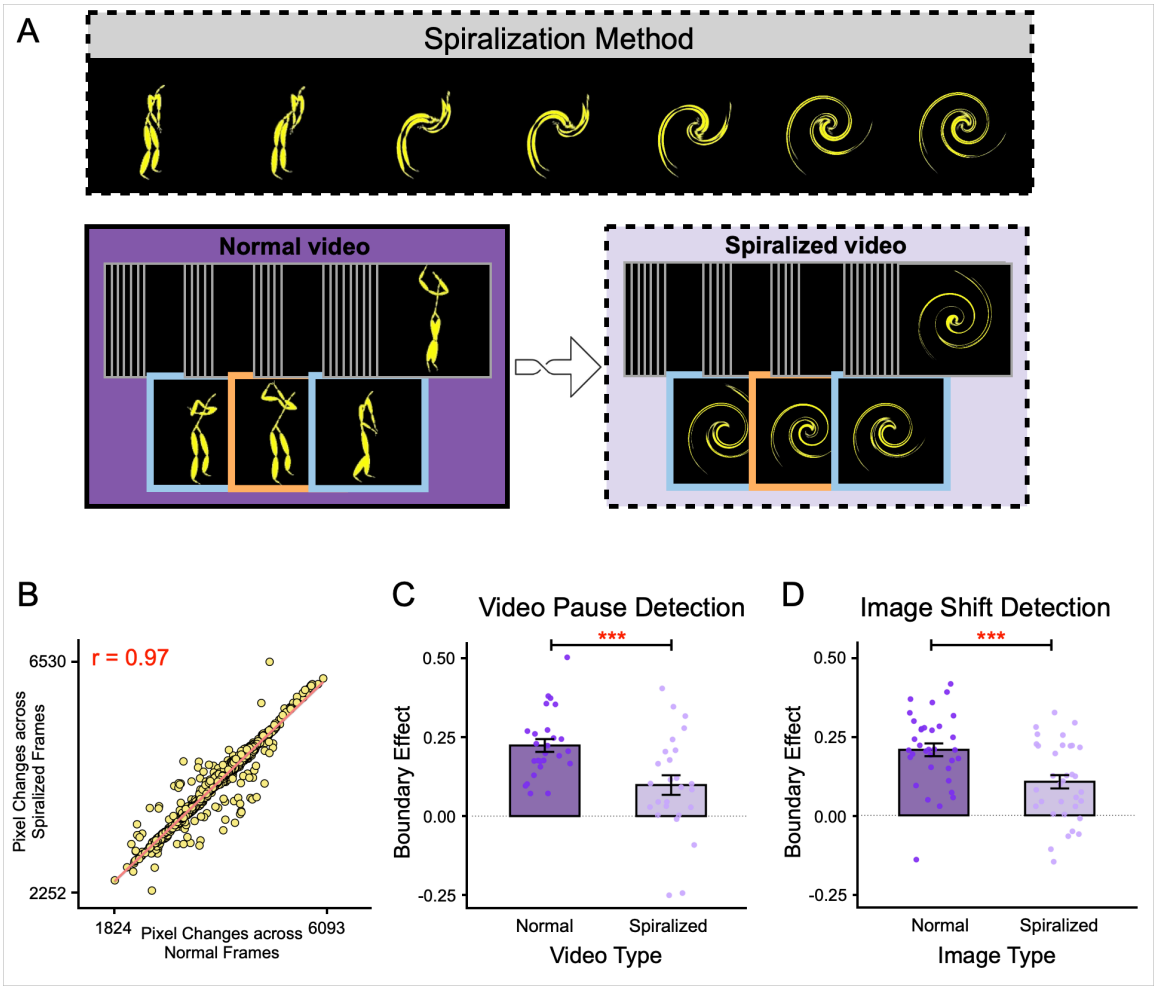

Figure 5: (A) Method used to generate the spiralized video stimuli. For each action video, we radially twisted all its frames using a spiral-shaped filter, rendering an abstract, meaningless pattern that maintained many of the lower-level spatiotemporal dynamics of the video clips. (B) We computed pixel-level changes between every pair of consecutive frames for both the normal videos and spiralized videos. We observed a high correlation between these pixel-level changes in normal versus spiralized videos. Experiments 4a and 4b used the same design as Experiment 2 and 3 respectively, except that observers saw the spiralized versions of videos and images. (C) In Experiment 4a, the boundary effect (i.e., the reduced detectability of pauses at boundaries relative to non-boundaries) was significantly weakened in the altered videos. (D) In Experiment 4b, the boundary effect decreased for the spiralized images. Error bars = 1 s.e.m. Asterisks indicate significant differences between means (***$p < .001$).

the previous experiment, yielding 20 (actions) × 3 (pre-, on-, post-boundary) × 2 (with/without pause) × 2 (normal, spiralized video) + 4 practice trials = 244 trials. We imposed brief pauses either at the pre-boundary, boundary, or post-boundary frames in both the normal and spiralized video conditions, using the same frame locations for each condition. Trial order was randomized across participants (except for practice trials).

5.2. Results

Eight participants were excluded for low accuracy ($< 57\%$, not significantly above chance according to a binominial test with $\alpha = .05$) and 1 for failing to provide complete data, leaving 26 participants with analyzable data, with a mean accuracy of 75.3%. Significant boundary effects were found in both types of videos: Observers were less accurate in detecting brief pauses at boundary frames relative to non-boundary frames (Normal stimuli: $t(25) = 11.05$, $p = 4.12 \times 10^{-11}$, $d = 2.17$, 95% $CI_{effect} = 0.22[0.18, 0.26]$; Spiralized: $t(25) = 3.16$, $p = 0.004$, $d = 0.62$, 95% $CI_{effect} = 0.10[0.041, 0.16]$). However, the boundary effect (i.e., the decreased detection accuracy at boundary relative to non-boundary frames) was significantly weakened for the spiralized stimuli ($t(25) = 4.19$, $p = 0.00030$, $d = 0.82$, $p = 0.004$, $d = 0.62$, 95% $CI_{difference} = 0.12[0.059, 0.18]$)[1] (Figure 5C). This suggests that the perception of event boundaries may not solely rely on low-level physical changes in stimuli, but also on internal representations of structure; in this case, the structure of familiar actions.

6. Experiment 4b: Low-level static features

To control for lower-level change of visual features involved in the image switch, we again used spiralized images in a task similar to Experiment 3, allowing us to compare the boundary effect in normal images versus their spiralized counterparts.

6.1. Method

6.1.1. Participants

As stated our pre-registration, 35 participants were recruited through Prolific.

[1]For Experiments 4a–b and 5a–b, in addition to the main analysis, we had also pre-registered a repeated-measures analysis of variance (ANOVA) as secondary analysis, which we reported in Supplemental Materials.

6.1.2. Stimuli and procedure

The experimental design was the same as Experiment 4a, except that two types of images were used as in Experiment 3: normal versus spiralized. The Image Type condition was crossed with other conditions as in previous experiments, yielding 20 (actions) × 3 (pre-, on-, post-boundary) × 2 (with, without pause) × 2 (normal, spiralized image) + 4 practice trials = 244 trials. Participants were instructed to detect whether the last image in the sequence was identical to the one immediately preceding it (Figure 4A). Trial order was randomized across participants (except for practice trials).

6.2. Results

One participant was excluded for failing to submit complete data, leaving 34 participants with analyzable data, with a mean accuracy of 72.4%. Thirty-four observers attempted to detect image changes across both conditions. We again observed a significant boundary effect in the normal images ($t(33) = 10.31$, $p = 7.52 \times 10^{-12}$, $d = 1.77$, 95% $CI_{effects} = 0.21[0.17, 0.25]$), and also observed this effect in the spiralized images ($t(33) = 5.04$, $p = 1.63 \times 10^{-5}$, $d = 0.86$, 95% $CI_{effects} = 0.11[0.065, 0.15]$); however, the effect was about half as strong in the spiralized images versus the normal images ($t(33) = 3.59$, $p = 0.0011$, $d = 0.62$, 95% $CI_{difference} = 0.10[0.045, 0.16]$, Figure 5D), echoing the results of Experiment 4a. This result suggests again that in addition to stimulus features, representations of internal structures in actions may be involved in rapid, spontaneous event segmentation in visual perception.

7. Experiment 5a: Point-light walkers

Our last two experiments aimed to replicate our findings in the case of biological motion. Although the twisting procedure used in Experiment 4a and 4b maintained pixel-level changes in the action videos, it inevitably altered other features, like motion direction and figure rigidity. In this experiment, we re-created the 20 actions as "point-light walkers" (PLWs)(Johansson, 1973; Van Boxtel & Lu, 2013) and attempted to replicate the previous effects. Thus, each action video now consisted of a number of coordinated moving points that represent the joints of the human figures. In this experiment, we presented PLWs either upright or upside-down. Inverted PLWs serve as an interesting case wherein the semantic information of actions is only partially preserved and can impede recognition (Shipley, 2003; Sumi, 1984). We asked whether the boundary effect emerged in visually minimal PLW stimuli, and whether the effect differed between upright versus inverted stimuli.

7.1. Method

7.1.1. Participants

Consistent with our pre-registration, 80 participants were recruited through Prolific. We increased the sample size for this experiment based on the preliminary result of a small pilot study.

7.1.2. Stimuli and Procedure

This experiment transformed the 20 original action videos into "point-light walker" stimuli. This was implemented using the BioMotion MATLAB toolbox (Van Boxtel & Lu, 2013). The number of joints used to define actions ranged from 20 to 30. This new set of videos depicted the same 20 actions as the ones used in previous experiments, yet were slightly different in the number of frames as a higher sampling rate was used. The boundary frame was hand-selected as the frame that had the matching posture as the boundary frame used in previous experiments, and the pre- and post-boundary frames were selected as before.

These PLW stimuli are presented either upright or upside-down. To reduce the number of repetitions of each action video, we asked participants to view each action only once in each trial and detect if the video was briefly frozen at any point. A 120-ms pause was applied to pre-, on-, and post-boundary frames of each action video. All 20 actions appeared for each of these trial types, for 20 (actions) $\times$ 3 (pre-, on-, post-boundary) $\times$ 2 (with, without pause) $\times$ 2 (upright, inverted video) + 4 practice trials = 244 trials. We note that because the experiments were crowd-sourced, we did not monitor participants for any behavior that could change the perceived orientation of the video clips (e.g., head-tilting).

7.2. Results

18 participants were excluded for low accuracy ($< 57\%$) and 1 participant for failing to submit a complete data set, leaving 61 participants with analyzable data, with a mean accuracy of 66.1%. The boundary effects arose similarly in both the upright videos ($t(60) = 13.34$, $p = 1.81 \times 10^{-19}$, $d = 1.72$, 95% $CI_{effects} = 0.24[0.21, 0.28]$) and inverted videos ($t(60) = 11.35$, $p = 1.82 \times 10^{-16}$, $d = 1.47$, 95% $CI_{effects} = 0.21[0.17, 0.25]$). The effects appear to be stronger in upright videos than inverted videos (24.4% versus 21.1%), though the difference between two conditions was marginally reliable given our pre-registered criteria ($t(60) = 1.75$, $p = 0.085$, $d = 0.22$, 95% $CI_{difference} = 0.033[-0.004, 0.07]$, Figure 6B). The subtle reduction of boundary effect in inverted actions suggest that 1) spatiotemproal information in the original PLWs, which is also preserved intact in their inverted forms, contributes to perceiving boundaries, and 2) disrupting the high-level semantic information may still weaken the perception of event boundaries.

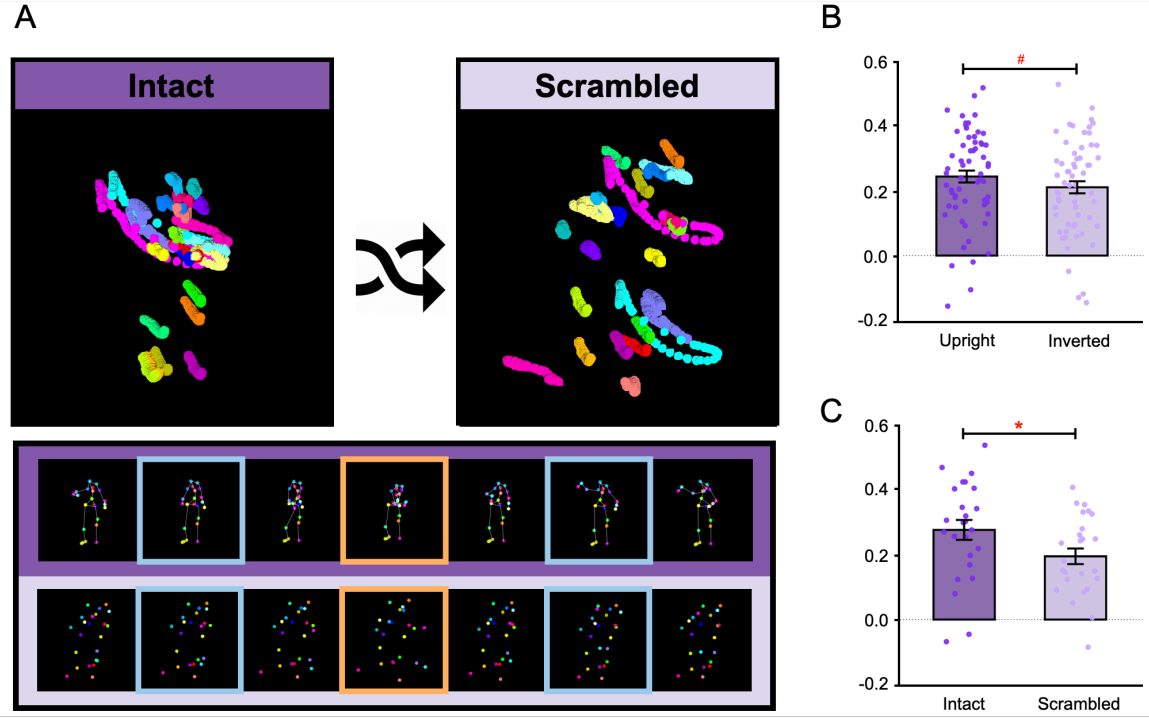

Figure 6: (A) All 20 original action videos were transformed into point-light walker stimuli, either in an intact form or a scrambled form (Van Boxtel & Lu, 2013). For a given action, the movements of all the individual points (joints) were identical between its intact form and scrambled form. Here, each color represents a unique joint, and the exact movement trajectory of each colored joint across all frames was identical between two types of stimuli. Similar to Experiment 2, a brief pause could occur at the pre-boundary, boundary, or post-boundary frame. (Note that all the visual points of the actual stimuli were colored yellow as in the original skeletonized stimuli, and the underlying skeleton, indicated by the thin lines in the above figure, was not visible to observers.) (B) In Experiment 5a, the reduction of boundary effect in upside-down PLWs was marginally significant relative to upright stimuli. (C) In Experiment 5b, the boundary effect was weaker when the global configuration of the observed agent was disrupted. Error bars = 1 s.e.m. Asterisks indicate significant differences between means for two-tailed t-test ($*p < .05$; $\#p < .1$)

8. Experiment 5b: Scrambled PLWs

Upside-down PLW actions are less recognizable yet still preserve partial semantic information (e.g., people may see still actions but misperceive some properties; See Barclay et al., 1978; Sumi, 1984). Is there a better way to eliminate semantic information in PLW stimuli? Here, we shuffled the initial positions of these "joints" such that the configuration of the body shape no longer resembled the human figure, while the local joint motion and rigidity was maintained. Thus, from the observer's perspective, the spatiotemporal motion perceived in the intact point-light actors was preserved in the scrambled videos, but the stimuli lost all semantic meaning.

8.1. Method

8.1.1. Participants

Consistent with our pre-registration, 35 participants were recruited through Prolific.

8.1.2. Stimuli

8.1.3. Procedure

These point-light videos used in Experiment 4a were made to "scrambled" versions, where the initial positions of the joints were randomly selected — separately by x, y and z — from the ranges of width, height, and depth of the original videos (Figure 6A). Readers can experience the similar motion in intact and scrambled videos at https://zk.actlabresearch.org/segmentation/plw.html

The experimental design was the same as Experiment 4a, except that: 1) point-light stimuli (normal and scrambled videos) were used, and 2) the pause duration was always 75 ms (to counter the overall increased difficulty of the task). Similar to Experiment 4a, all 20 actions appeared for each of these trial types, for 20 (actions) $\times$ 3 (pre-, on-, post-boundary) $\times$ 2 (with, without pause) $\times$ 2 (normal, scrambled video) + 4 practice trials = 244 trials. Trial order was randomized across participants (except for practice trials).

8.2. Results

Ten participants were excluded for low accuracy ($< 57\%$), leaving 25 participants with analyzable data, with a mean accuracy of 76.6%. We again observed robust boundary effects in both the intact ($t(24) = 9.02$, $p = 3.53 \times 10^{-9}$, $d = 1.80$, 95% $CI_{effects} = 0.28[0.22, 0.33]$) and scrambled videos ($t(24) = 8.11$, $p = 2.45 \times 10^{-8}$, $d = 1.62$, 95% $CI_{effects} = 0.19[0.15, 0.24]$). Critically, these boundary effects were reduced in the scrambled videos, where semantic action structure was not perceivable

($t(24) = 2.22$, $p = 0.036$, $d = 0.44$, 95% $CI_{difference} = 0.081[0.01, 0.15]$, Figure 6C). Thus, while we again observed a significant contribution of lower-level spatiotemporal dynamics to the visual segmentation of actions (i.e., significant boundary effects in the scrambled condition), we also again showed that higher-level representations of actions may also play a significant role in the perception of action event structure.

9. Discussion

Here we asked if event structure in observed actions are represented by the visual system in a manner that might go beyond their low-level visual features alone. Using motion-captured videos, static action images, and biological motion stimuli, we consistently observed that the transition of an action's salient internal steps impaired observers' ability to visually detect subtle changes in the stimuli. Control studies further suggested that this boundary representation is not solely driven by basic visual features of the stimulus that happen to change at the boundary (e.g., spatiotemporal features and dynamics), but also by internal models of the actions themselves (Experiments 4a–b and 5a–b). That is, high-level event structure of actions may be represented in visual perception, and thus may support segmentation.

A large body of literature has explored event segmentation *between* fully distinct actions, which typically involves large salient changes in the visual scene or the location and movements of an observed agent and surrounding objects (Baldwin et al., 2008; Buchsbaum et al., 2015; Franklin et al., 2020; Lea et al., 2016; Newtson, 1973; Newtson et al., 1977; Pomp et al., 2024; Swallow et al., 2009; Wang et al., 2013; Zacks et al., 2009). Our study was designed to focus on more subtle event boundaries that occur *within* short, continuous, bounded actions (Y. Ji & Papafragou, 2022; Vendler, 1957). The action segmentation in our study was thus more rapid and finer-grained than those described as "fine" units (with median lengths of 10–15 s) in previous work (Hard et al., 2011; Yates et al., 2024; Zacks & Tversky, 2001; Zacks et al., 2009). Though finer boundaries are thought to be more perceptually determined and driven by low-level changes of pixel-level movement and motion (Hard et al., 2006; Papeo et al., 2024; Zacks et al., 2009), here we found a comparable contribution of internal representations of the structure of actions in visual event perception.

This work builds on a number of recent studies that have discussed the spontaneous nature of representing event boundaries in visual perception. For example, in one recent study observers were less sensitive to visual interruptions at the end points of events (e.g. a girl folding her handkerchief) compared to the middle points(Y. Ji & Papafragou, 2022). Additionally, the boundary of simple physical events (such as

collision, containment, or falling) has been found to impede the processing of irrelevant information, but facilitate relevant information (Huff et al., 2012; Yates et al., 2022). Such results have been explained by the rise of prediction error at the boundary and subsequent enhanced attention to event information (Pradhan & Kumar, 2022; Reynolds et al., 2007; Zacks et al., 2007, 2011). Nevertheless, in these cases, feature, motion, and movement information was inevitably correlated with the event structure of ongoing visual input. Our study contributes to this thread of work by attempting to dissociate two possible mechanisms that may account for the perceptual consequences of event segmentation: lower-level spatiotemporal dynamics in the stimulus, and higher-level internal representations of the structure of events. In that vein, our findings suggest that predictive processes that bridge discrete events during visual perception may employ both low-level information and high-level information to generate predictions.

Our study built on the logic of previous studies, focusing on the perception of disruptions or distractions at event boundaries (Huff et al., 2012; Repp, 1992; Yates et al., 2024). However, the interpretation of these detection effects as reflections of attentional or predictive processes could be better fleshed out; future work could look for other perceptual effects at boundaries in observed actions, such as increased processing of certain features versus others (Baker & Levin, 2015; Yates et al., 2024), and temporal distortions (Goh et al., in press; Ongchoco et al., 2023a). Moreover, our study employed salient, simple, familiar actions; this was by design, to make the semantic structure overdetermined. In future work, we could ask how more novel or unfamiliar actions are segmented, or how learning about new actions may affect the perception of event boundaries. Lastly, how our findings connect to the learning of actions themselves could provide insights into why we segment actions the way we do. For example, in learning a new perceptuo-motor skill, like a tennis serve, coaches and teachers tend to divide an otherwise smooth, rapid movement into multiple discrete "chunks" based on an internal structure (Fitts, 1964; Sakai et al., 2003).

Research in human cognition (Hard et al., 2011; Newtson et al., 1977; Zacks et al., 2009) and computer vision(Lea et al., 2016; Wang et al., 2013) both suggest an important role of motion and spatiotemporal information in action recognition and segmentation. For the subtle event boundaries like the ones we studied here (i.e., boundaries occurring over short time-scales in single, continuous actions), visual motion certainly still played a major role in segmentation (Zacks et al., 2009). Indeed, our results using multiple forms of control stimuli consistently pointed to robust effects of lower-level features. However, we also found that action-specific semantic information appeared to contribute to visual event segmentation, suggesting that the visual system spontaneously represents the high-level temporal structure

of actions. This finding arguably fits with the recent resurgence of interest in ideas like the Language-of-Thought approach to vision, which proposes multiple types of compositional and structured representations in perception(Hafri et al., 2024; Papeo et al., 2024; Quilty-Dunn et al., 2023). The current study presents novel evidence supporting the idea that the mind spontaneously represents the flow of experiences as discrete, inherently structured events unfolding over time.

Acknowledgment

This work was supported by grant R01 NS134754 (S.D.M.) from the National Institutes of Health.

References

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron*, *109*(2), 377–390.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition*, *53*(3), 239–261.

Ayzenberg, V., & Lourenco, S. (2022). Perception of an object's global shape is best described by a model of skeletal structure in human infants. *elife*, *11*, e74943.

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & psychophysics*, *23*, 145–152.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive psychology*, *76*, 30–77.

Cohn, N., Holcomb, P., Jackendoff, R., & Kuperberg, G. (2012). Segmenting visual narratives: Evidence for constituent structure in comics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34).

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (pea) nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive psychology*, *65*(1), 1–38.

Comrie, B. (1976). Aspect: An introduction to the study of verbal aspect and related problems. *Cambridge UP*.

Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta psychologica*, *149*, 69–77.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277.

Ezzyat, Y., & Clements, A. (2024). Neural activity differentiates novel and learned event boundaries. *Journal of Neuroscience*, *44*(38).

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Fitts, P. M. (1964). Perceptual-motor skill learning. In *Categories of human learning* (pp. 243–285). Elsevier.

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured event memory: A neuro-symbolic model of event cognition. *Psychological review*, *127*(3), 327.

Goh, R. Z., Zhou, H., Firestone, C., & Phillips, I. (in press). Event-based warping: A relative distortion of time within events. *Journal of Experimental Psychology: General*.

Hafri, A., Bonner, M. F., Landau, B., & Firestone, C. (2024). A phone in a basket looks like a knife in a cup: Role-filler independence in visual processing. *Open Mind*, *8*, 766–794.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52.

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & cognition*, *34*(6), 1221–1235.

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of experimental psychology: General*, *140*(4), 586.

Hemeren, P. E., & Thill, S. (2011). Deriving motor primitives through action segmentation. *Frontiers in psychology*, *1*, 243.

Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*(7), 848–864.

Ji, H., & Scholl, B. J. (2024). "visual verbs": Dynamic event types are extracted spontaneously during visual perception. *Journal of Experimental Psychology: General*, *153*(10), 2441.

Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language*, *127*, 104353.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, *14*, 201–211.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences*, *12*(2), 72–79.

Lea, C., Reiter, A., Vidal, R., & Hager, G. D. (2016). Segmental spatiotemporal cnns for fine-grained action segmentation. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, 36–52.

Lee, H., & Chen, J. (2022). A generalized cortical activity pattern at internally generated mental context boundaries during unguided narrative recall. *Elife*, *11*, e73693.

Lovett, A., & Franconeri, S. L. (2017). Topological relations between objects are categorically coded. *Psychological science*, *28*(10), 1408–1418.

Mittwoch, A. (2013). On the criteria for distinguishing accomplishments from activities, and two types of aspectual misfits. In *Studies in the composition and decomposition of event predicates* (pp. 27–48). Springer.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of personality and social psychology*, *28*(1), 28.

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and social psychology*, *35*(12), 847.

Ongchoco, J. D. K., & Scholl, B. J. (2019). Did that just happen? event segmentation influences enumeration and working memory for simple overlapping visual events. *Cognition*, *187*, 188–197.

Ongchoco, J. D. K., Walter-Terrill, R., & Scholl, B. J. (2023b). Visual event boundaries restrict anchoring effects in decision-making. *Proceedings of the National Academy of Sciences*, *120*(44), e2303883120.

Ongchoco, J. D. K., Yates, T. S., & Scholl, B. J. (2023a). Event segmentation structures temporal experience: Simultaneous dilation and contraction in rhythmic reproductions. *Journal of Experimental Psychology: General*.

Papeo, L., Vettori, S., Serraille, E., Odin, C., Rostami, F., & Hochmann, J.-R. (2024). Abstract thematic roles in infants' representation of social events. *Current Biology*, *34*(18), 4294–4300.

Pomp, J., Garlichs, A., Kulvicius, T., Tamosiunaite, M., Wurm, M. F., Zahedi, A., Wörgötter, F., & Schubotz, R. I. (2024). Action segmentation in the

brain: The role of object–action associations. *Journal of cognitive neuroscience*, *36*(9), 1784–1806.
Pradhan, R., & Kumar, D. (2022). Event segmentation and event boundary advantage: Role of attention and postencoding processing. *Journal of Experimental Psychology: General*, *151*(7), 1542.
Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, *46*, e261.
Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, *21*(4), 269–272.
Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & cognition*, *34*, 1150–1156.
Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, *44*(3), 241–281.
Repp, B. H. (1998). Variations on a theme by chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 791.
Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive science*, *31*(4), 613–643.
Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental brain research*, *152*, 229–242.
Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature neuroscience*, *16*(4), 486–492.
Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological science*, *14*(4), 377–380.
Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 335–345.
Strickland, B., & Scholl, B. J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General*, *144*(3), 570.
Sumi, S. (1984). Upside-down presentation of the johansson moving light-spot pattern. *Perception*, *13*(3), 283–286.
Sun, Z., & Firestone, C. (2021). Curious objects: How visual complexity guides attention and engagement. *Cognitive Science*, *45*(4), e12933.

Sun, Z., Firestone, C., & Hafri, A. (2025). The psychophysics of compositionality: Relational scene perception occurs in a canonical order. *psyArxiv preprint*.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236.

Tauzin, T. (2015). Simple visual cues of event boundaries. *Acta psychologica*, *158*, 8–18.

Van Boxtel, J. J., & Lu, H. (2013). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *Journal of vision*, *13*(12), 7–7.

Vendler, Z. (1957). Verbs and times. *The philosophical review*, *66*(2), 143–160.

Wang, H., Kläser, A., Schmid, C., & Liu, C.-L. (2013). Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision*, *103*, 60–79.

Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., Baldassano, C., & Turk-Browne, N. B. (2022). Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences*, *119*(43), e2200257119.

Yates, T. S., Yasuda, S., & Yildirim, I. (2024). Temporal segmentation and "look ahead" simulation: Physical events structure visual perception of intuitive physics. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(8), 859.

Zacks, J. M. (2020). Event perception and memory. *Annual review of psychology*, *71*(1), 165–191.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature neuroscience*, *4*(6), 651–655.

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201–216.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of cognitive neuroscience*, *23*(12), 4057–4066.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological bulletin*, *133*(2), 273.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain research*, *1076*(1), 150–162.

669 Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception.
670 *Psychological bulletin*, *127*(1), 3.
671 Zheng, Y., Zacks, J. M., & Markson, L. (2020). The development of event perception
672 and memory. *Cognitive Development*, *54*, 100848.

Supplement

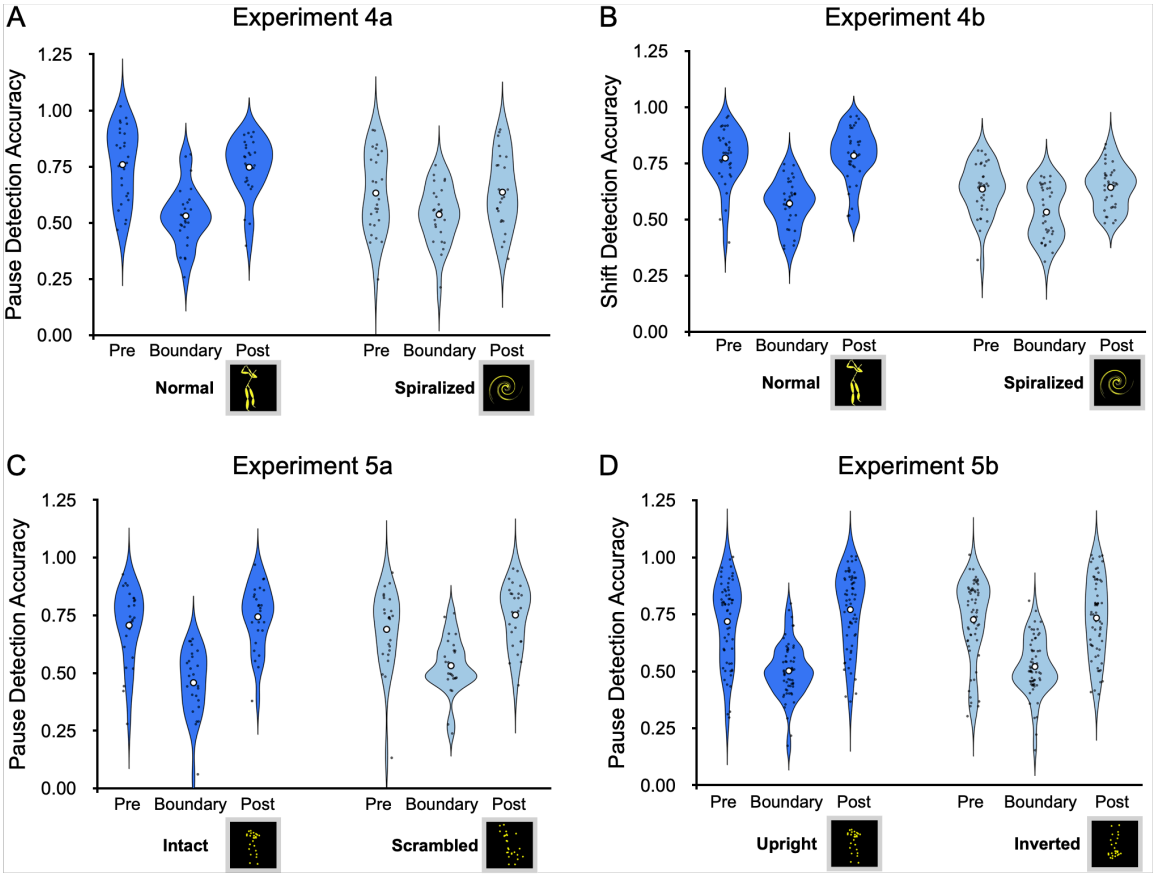
A
Experiment 4a
Pause Detection Accuracy
1.25
1.00
0.75
0.50
0.25
0.00
Pre Boundary Post
Pre Boundary Post
Normal
Spiralized
B
Experiment 4b
Shift Detection Accuracy
1.25
1.00
0.75
0.50
0.25
0.00
Pre Boundary Post
Pre Boundary Post
Normal
Spiralized
C
Experiment 5a
Pause Detection Accuracy
1.25
1.00
0.75
0.50
0.25
0.00
Pre Boundary Post
Pre Boundary Post
Intact
Scrambled
D
Experiment 5b
Pause Detection Accuracy
1.25
1.00
0.75
0.50
0.25
0.00
Pre Boundary Post
Pre Boundary Post
Upright
Inverted

Figure S1: In Experiments 4a, 4b, 5a and 5b, the tested frame and stimulus type were within-subject factors. In addition to the main analysis, we had also pre-registered a repeated-measures analysis of variance (ANOVA) across subject means of detection accuracy, as a secondary analysis. Note that we determined sample size only using power analyses on the main results (i.e., the pre-registered paired t-tests on the boundary effect) using pilot studies. Thus, the ANOVAs here likely do not have sufficient power with the current sample size. (A) In Experiment 4a, a 2 (video type: normal vs. spiralized) $\times$ 3 (tested frame: non-boundary vs. boundary) repeated-measures ANOVA showed a significant main effect for Tested Frame, $F(1, 96) = 9.65$, $p = 0.0025$; and a marginally significant effect Video Type, $F(1, 96) = 2.69$, $p = 0.10$ (No significant interaction: $F(1, 96) = 1.37$, $p = 0.24$). (B) In Experiment 4b, we found a significant main effect for both Tested Frame ($F(1, 128) = 17.87$, $p = 4.47 \times 10^{-5}$) and Image Type ($F(1, 128) = 7.38$, $p = 0.0075$), as well as a marginal significant interaction between two factors ($F(1, 128) = 3.75$, $p = 0.055$). (C) The results of Experiment 5a only revealed a significant main effect for tested frame, $F(1, 92) = 36.47$, $p = 3.24 \times 10^{-8}$ (Video Type: $F(1, 92) = 1.66$, $p = 0.20$; no significant interaction: $F(1, 92) = 0.39$, $p = 0.53$). (D) The results of Experiment 5b revealed a significant main effect for tested frame, $F(1, 236) = 53.88$, $p = 3.42 \times 10^{-12}$ (Video Type: $F(1, 236) = 0.003$, $p = 0.96$; no significant interaction: $F(1, 236) = 0.27$, $p = 0.60$).