Viewers Spontaneously Represent Event Temporal Structure

Yue Ji (jiyue@bit.edu.cn)
School of Foreign Languages, Beijing Institute of Technology. No. 5 South Street, Zhongguancun
Beijing, 100081 China

Anna Papafragou (anna4@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania. 3401-C Walnut St.
Philadelphia, PA, 19104 USA

Abstract

Events are considered as temporal segments with a beginning and an endpoint. A large body of philosophical and linguistic literature on events distinguishes between *bounded* events that are composed of distinct temporal stages leading to culmination (e.g., fix a car) and *unbounded* events that are composed of largely undifferentiated stages and lack an inherent endpoint (e.g., drive a car). In the present study, we show that event viewers spontaneously compute this distinction through an interruption detection task. People watched videos of either bounded or unbounded events that included a visual interruption lasting .03s placed at either the midpoint or close to the endpoint of the event stimulus. People had to indicate whether they saw an interruption after watching each video (Experiments 1) or respond as soon as they detected an interruption while watching each video (Experiment 2). In both cases, the endpoint-midpoint difference depended on whether participants were watching an event that was bounded or unbounded. This result suggests that, as people perceive dynamic events, they spontaneously track boundedness, or the internal temporal structure of events.

Keywords: boundedness; aspect; event structure; event perception

Introduction

The world is a continuous flow of activity, but we segment our continuous experience in terms of concrete units with beginnings and ends, i.e., events. According to a prominent account (Event Segmentation Theory, or EST; Zacks et al., 2007), the process of segmenting events is guided by stable working memory representations, known as *event models*. Event models contain some structured information about events (including event participants, their intentions and goals, as well as temporal, spatial, and causal relations among event participants; see Radvansky & Zacks, 2014). Event models help observers make predictions about upcoming happenings. The perception of event boundaries depends on these predictions: when important situation features change, people cannot accurately predict what is coming next and have to update their event models. The moment when maximal prediction errors occur is thus experienced as an event boundary.

A key finding from the literature on event segmentation is that event boundaries are influential for event processing. For instance, visual stimuli that include only event boundaries are understood and recalled better than stimuli that include only event middles (Newston & Engquist, 1976; Schwan & Garsofsky, 2004). Similarly, objects relevant to an event boundary are recognized more easily than objects relevant to non-boundary moments (Swallow et al., 2009), and objects external to the event stimulus are detected more accurately when inserted outside of event boundaries (Huff et al., 2012). A plausible explanation for the advantage of event boundaries is offered by EST: once an event comes to an end, a range of possible new events may follow; the transition is less predictable and thus requires more processing resources (Zacks et al., 2007). In support of this idea, people spend more time at event boundaries when reading event descriptions or watching slideshows of events at their own pace (Hard et al., 2011; Pettijohn & Radvansky, 2016). In this line of reasoning, attention is organized in line with event segmental structure, with more attention being allocated to the less predictable event boundaries. This attentional bias may lead to the privileged status of event boundaries in comprehension and memory.

Despite the emphasis on how people identify event boundaries within the above literature, a topic that has received much less discussion is how people process the representational unit *within* event boundaries (see Huff & Papenmeier, 2017). Typically, the research on event segmentation identifies an event as "a segment of time at a given location that is conceived by an observer to have a beginning and an end" (Zacks & Tversky, 2001) but does not address how people represent the content of specific events. Here we propose that, to better understand how the human mind represents events, we need to consider the temporal texture *within* individual events and event classes.

According to a long linguistic and philosophical study of events (see Filip, 2012; van Hout, 2016), language describes a situation as either a *bounded* or an *unbounded* event. The two types of events have different internal structures and different ways in which they come to an end. For instance, the sentence "A girl fixed a car" encodes an experience as a bounded event: this event has a non-homogenous structure consisting of distinct, articulated stages (e.g., opening the car hood, checking the engine, etc.) that lead to a "built-in terminal point" (Comrie, 1976), "climax" (Vendler, 1957) or "culmination" (Parsons, 1990) - the moment when the car starts to work again. The endpoint of bounded events is projected "from the outset" and is naturally achieved unless there is an interruption (Mittwoch, 2013). By contrast, the sentence "A girl drove a car" encodes an experience as an unbounded event: this event has a homogenous structure that lacks distinct stages since “any part of the process is of the same nature as the whole” (Vendler, 1957) - each moment of the girl’s action can still be described as an event of driving a car. Unbounded events have no specified endpoint and may end at an arbitrary moment (in the example above, the endpoint could be any moment when the girl stops driving).

Recent experimental work reports that viewers extract boundedness information when processing naturalistic visual events, even when they are not engaged in the process of producing or comprehending event descriptions. In a direct demonstration (Ji & Papafragou, 2020), participants watched videos of a character perform everyday actions; some videos were marked by a red frame in a way that corresponded to either the bounded or the unbounded event category. The participants succeeded in identifying whether the red frame applied to a new set of events. Other studies have offered evidence that boundedness cross-cuts linguistic and visual stimuli (e.g., Malaia et al., 2012; Strickland et al., 2015; Wehry et al., 2019; Wellwood et al., 2018).

How exactly does boundedness contribute to conceptual event representations? A first possibility is that boundedness is computed as part of the continually evolving event representation that viewers generate spontaneously as they process dynamic visual input. [1] On this hypothesis, boundedness could be captured by extending the mechanisms outlined in Event Segmentation Theory (Swallow et al., 2009; Zacks et al., 2007). On this theory, viewers predict what is going to happen next in the perceptual stream, and update their working model of an event continuously. Boundedness can be viewed as an outcome of viewers’ sensitivity to accumulating change *within* the boundaries of an event, even when the change does not warrant inserting an event breakpoint. During unbounded (homogeneous) events, observers can easily predict what comes next based on what is happening in the moment, and treat temporal slices of the event similarly since they are equally predictable. By contrast, during bounded (non-homogeneous) events, different temporal slices represent different stages of development, with the moment of the event endpoint or culmination being the least predictable.

According to an alternative hypothesis, however, awareness of bounded/unbounded event classes might arise through explicit and deliberate observation of commonalities among event exemplars but does not drive event apprehension itself. In other words, boundedness can be computed by viewers as an abstraction over events but does not emerge during ordinary event processing. Notice that the tasks used to probe non-linguistic boundedness have typically been explicit and involved intentionally inspecting specific event tokens for the purposes of forming an event class (e.g., Ji & Papafragou, 2020). To settle this issue in favor of the spontaneity hypothesis, one would need evidence that observers compute event boundedness as they process naturalistic events even when they are engaged in some orthogonal task.

In the present study, we hypothesized that event viewers spontaneously track the temporal texture of bounded and unbounded events. To test this hypothesis, we introduced very brief disruptions at different time points within videos of bounded vs. unbounded events during which the visual stimulus became blurry. The observers’ task was to detect these disruptions. Observers had to respond either after watching a video (Experiments 1) or as soon as they detected the interruption while they watched the video (Experiment 2). The disruptions were inserted as an attentional probe and thus detection accuracy should be lower and response times should be longer when more processing resources were drawn by the event stimuli (see also Huff et al., 2012). If boundedness is computed as part of event apprehension, we should observe differential sensitivity to the placement of visual interruptions depending on the boundedness of the stimulus. Specifically, for bounded events whose internal texture has distinct sub-stages and leads to the highly informative moment of culmination, disruptions should be harder to detect when they appear close to the event endpoint compared to the midpoint. By contrast, for unbounded events whose temporal texture is largely undifferentiated, there should be little or no difference in detection of disruptions placed at midpoints vs. endpoints of event stimuli.

Experiment 1

Method

Participants Sixty-four adults (age range: 18-23) participated in the experiment. Our sample size was decided based on the sample size in similar studies on event perception and memory (e.g., Huff et al., 2012; Papenmeier, et al., 2019). All participants were undergraduates at a major university on the East Coast of the US. Data from 3 additional adults were collected but excluded because they kept giving *Yes* responses throughout the experiment.

Stimuli We used the same 20 pairs of videos as Ji and Papafragou (2020). Paired videos showed a bounded and an unbounded event, and had the same duration (4-12s, M = 6.7s; see Table 1). All of the events involved the same girl who did a familiar everyday action in a sparse room. The action began with the girl picking up an object or tool from a tabletop surface and came to an end with her putting down the object or tool and removing her hands from the table. As in the linguistic literature, the contrast between bounded and unbounded events was due to either the nature of the action or the nature of the affected object (see Tenny, 1987). For half of the videos, paired bounded and unbounded events

[1] Spontaneous cognitive processes are unconscious and involuntary, even though their operation is determined by attention or some other form of calibration (Carruthers, 2017; O’Grady, et al., 2020). As such, they differ from automatic processes that are reflexive and cannot be inhibited (ibid.).

Table 1: Stimulus events in Experiment 1

Phase	Boundedness Source	No.	Bounded Events	Unbounded Events
Practice	Nature of Action	1	close a fan	use a fan for oneself
		2	crack an egg	beat an egg
	Nature of Affected Object	3	cut a ribbon in half	cut ribbon from a roll
		4	stick a sticker	stick stickers
Testing	Nature of Action	5	fold up a handkerchief	wave a handkerchief
		6	put up one’s hair	scratch one’s hair
		7	stack a deck of cards	shuffle a deck of cards
		8	group pawns based on color	mix pawns of two colors
		9	dress a teddy bear	pat a teddy bear
		10	roll up a towel	twist a towel
		11	fill a glass with milk	shake a bottle of milk
		12	scoop up yogurt	stir yogurt
	Nature of Affected Object	13	draw a balloon	draw circles
		14	tie a knot	tie knots
		15	eat a pretzel	eat cheerios
		16	flip a postcard	flip pages
		17	peel a banana	crack peanuts
		18	blow a balloon	blow bubbles
		19	tear a paper towel	tear paper towels
		20	paint a star	paint stuff

involved the same object but differed in terms of the nature of the action performed on the object: the bounded event displayed an action that caused a clear and temporally demarcated change of state in the object (e.g., stack a deck of cards) while its unbounded counterpart did not involve such a change (e.g., shuffle a deck of cards). For the other half of the videos, the bounded and unbounded events involved the same action but differed in terms of the nature of the affected object: the bounded event involved a single object (e.g., blow a balloon) but its unbounded counterpart involved either an unspecified plurality of objects or a mass quantity (e.g., blow bubbles).

Two norming studies were conducted to ensure that viewers talked about and considered our stimuli as either a bounded or an unbounded event as expected. First, in an event description task, these videos successfully aligned with the linguistic boundedness distinction in English: stimuli of bounded events elicited bounded descriptions that included change-of-state predicates (e.g., stack a deck of cards) or quantified count noun phrases (e.g., blow a balloon) 98.2% of the time. Stimuli of unbounded events elicited unbounded verb phrases that included verbs of activity (e.g., shuffle a deck of cards) or unquantified noun phrases (bare plurals or mass nouns: e.g., blow bubbles) 92.8% of the time. Second, in a task that elicited judgment about the temporal structure of the stimuli, videos of bounded events were considered as “something with a beginning, midpoint and specific endpoint” 87.2% of the time while videos of unbounded events were considered as such only 20.3% of the time.

The videos were then edited in Corel VideoStudio X9 to introduce a “break” of 0.03s (i.e., 1 editing frame, with a video display rate of 30 frames per second; see also Hard et al., 2011; Strickland & Keil, 2011). The break consisted of a blurry picture created by applying an Iris Blur Effect in Adobe Photoshop CS 6 to portions of the original video (see the examples in Figure 1 and Figure 2). Each video was edited twice. In the mid-break version, the break replaced the frame that showed the temporal midpoint of the event (e.g., in the video of blowing a balloon with 300 frames, the mid-break replaced the 151st frame). In the late-break version, the break began at the point that corresponded to 80% of the event (e.g., in the same video of blowing a balloon, the late-break replaced the 241st frame). Edited videos were used as test items, and their original versions were used as fillers.

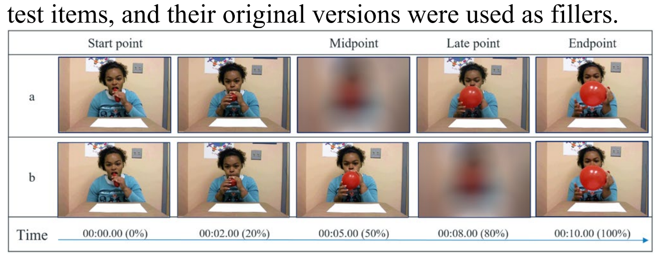

Figure 1: Examples of two versions of a bounded event (blow a balloon) in Experiment 1: (a) mid-break (b) late-break.

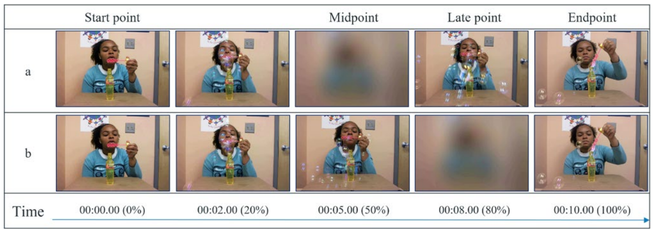

Figure 2: Examples of two versions of an unbounded event (blow bubbles) in Experiment 1: (a) mid-break (b) late-break.

The video stimuli of bounded events were arranged into 4 lists. Each list began with a practice phase composed of 4 videos. For this phase, the first and third videos always had a mid-break and a late-break respectively and the other two videos did not include a break. The same 4 events were used as practice items for all 4 lists but each event appeared in the mid-break version in one list, in the late-break version in a second list, and as a filler without any break in the remaining two lists. Within each list, the testing phase was composed of 8 test videos (4 with a mid-break, 4 with a late-break) and 8 fillers. Whether an event appeared as a test item or a filler was rotated across the lists. Unlike the practice phase, the events were presented in the same order across the 4 lists. Therefore, the order between test items and fillers differed among the lists. In each list, test items and fillers were intermixed such that items of the same type could not appear successively more than 3 times. The position of the break (mid vs. late) and the source of boundedness (action vs. affected object) in test videos were counterbalanced. The stimuli of unbounded events were also arranged into 4 lists in the same way.

Procedure Participants were randomly assigned to one of two conditions depending on the event type (Bounded or Unbounded) that they were exposed to throughout the experiment. Within each condition, they were randomly assigned to one of the 4 lists. Participants were told to watch each video carefully and decide whether they saw a break in the video. Responses were given by circling either "Break", or "No break" on an answer sheet. Participants were then given a practice phase meant to illustrate what a break was. After each practice trial, participants noted their answer, and then the experimenter gave the correct answer. If participants were wrong, the video was played a second time. In the testing phase, no feedback was given.

Results

"Break" responses to test items and "No break" responses to fillers were coded as correct. Overall, the accuracy of responses to fillers did not differ significantly between the Bounded (M = 93.8%) and Unbounded condition (M = 92.2%) ($t(62)$ = -.611, p = .543). Turning to test items (see Figure 3), we submitted the binary accuracy data to a mixed logit model with fixed effects of Event Type (Bounded vs. Unbounded), Break Placement (Mid vs. Late) and their interaction. All of the factors were coded using centered contrast (-0.5, 0.5). Random intercepts were provided for each Subject and each Item (Baayen, et al., 2008; Barr, 2008). [2] The analysis showed that the difference between Bounded (M = 87.5%) and Unbounded event types (M = 94.5%) approached significance (β =0.69, z = 1.80, p =.072). Similarly, there was a trend towards better break detection at midpoints (M = 94.5%) compared to late points (M = 87.5%) (β = -0.69, z = -1.87, p = .061). Crucially, there was a significant interaction between Event Type and Break Placement (β = 1.99, z = 2.70, p =.007). Participants watching videos of bounded events were better at detecting mid-breaks (M = 95.3%) than late-breaks (M = 79.7%) (β = -1.74, z = -3.53, p < .001). By contrast, participants watching videos of unbounded events did not differ in their detection of mid-breaks (M = 93.8%) and late-breaks (M = 95.3%, β = 0.31, z = 0.55, p = .581).

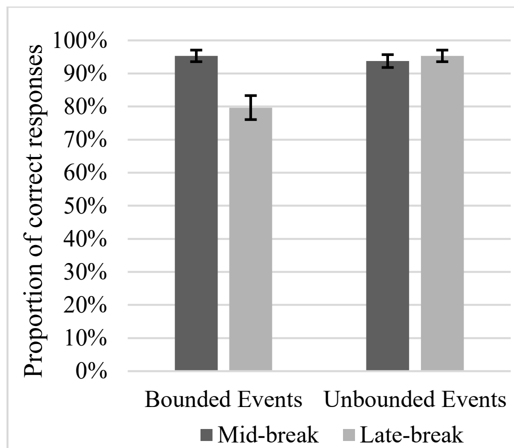

Figure 3: Proportion of correct responses in Experiment 1. Error bars represent ±SEM.

Discussion

In this experiment, viewers were more likely to miss a visual disruption of an event stimulus when the disruption occurred close to the event ending compared to the event midpoint, but only when perceiving a bounded event; there was no effect of the placement of the disruption when viewers perceived an unbounded event. This effect of event type on the detection of mid- and late-disruptions emerged even though neither the placement of the disruption nor the content of the disrupted event were relevant to the viewers' task. Therefore, the results support our hypothesis that viewers track the temporal structure of events as part of their event understanding.

[2] Adding Boundedness Source (Action vs. Affected Object) and List to the model did not reliably improve model fit so we excluded these factors from further analysis.

Experiment 2

In Experiments 1, participants gave a response after watching each video, and it remains possible that their detection of breaks was influenced by their construal of the whole event. To exclude this possibility, in Experiment 2, participants were asked to indicate detection of a break as soon as possible as they watched each video. If the effect of break placement in bounded but not unbounded events persists, it would strongly support the hypothesis that observers spontaneously track event boundedness during event perception.

Method

Participants Sixty-four adults (age range: 18-23) recruited from the undergraduate population of a major university on the East Coast of the US participated in the experiment for course credit. Data from 6 additional adults were collected but excluded: two participants did not understand the task; two participants always responded *Yes* throughout the experiment; one participant tended to respond multiple times in each trial during the experiment; one participant in the Bounded condition had an average response time more than 2 standard deviations above the average of participants in the same condition.

Stimuli Video stimuli were identical to those used in Experiment 1.

Procedure Experiment 2 was an online study conducted on the PennController platform for Internet Based Experiments (PCIbex, Zehr & Schwarz, 2018). Participants logged in to the experiment from their computer. Initial instructions informed them that they would watch some videos and that some of these videos contained a break. Their task was to detect the break as soon as they could while watching a video. They were told to press the spacebar immediately if they detected a break in the video, or press N at the end of the video if they did not see any break. In each trial, both response type and response time were recorded. If a participant did not respond within 5 seconds after the end of the video, the program automatically moved on to the next trial. As in Experiments 1, participants had a practice session to understand what a break was. During practice, participants received feedback on their response in each trial. At test, no feedback was given.

Results

We coded *Yes* responses (i.e., pressing the spacebar) to test items and *No* responses (i.e., pressing N) to fillers as correct. Errors included failure to detect the break in test items, false alarms and timeouts (N = 9, 0.9% of total responses). We further checked the response times in correct responses and recoded as errors any *Yes* responses that occurred before the time of the break in test videos (N = 40, 3.9% of total responses) and any *No* responses that occurred before the end of filler videos (N = 4, 0.4% of total responses).

Performance on filler items did not differ between event types (M = 80.5% for Bounded vs. 78.9% for the Unbounded events, $t(62) = .350$, $p > .250$). For test items, the same coding and analytic strategy was used as in Experiments 1. As shown in Figure 4, there was a significant effect of Break Placement, such that participants were better at detecting breaks at midpoints (M = 89.5%) than breaks close to event endpoints (M = 81.3%) ($\beta = -.65$, $z = -2.36$, $p = .018$). The difference between Bounded (M = 82.8%) and Unbounded (M = 87.9%) event types was not significant ($p = .27$), nor was there a significant interaction between Event Type and Break Placement ($p = .38$). [3]

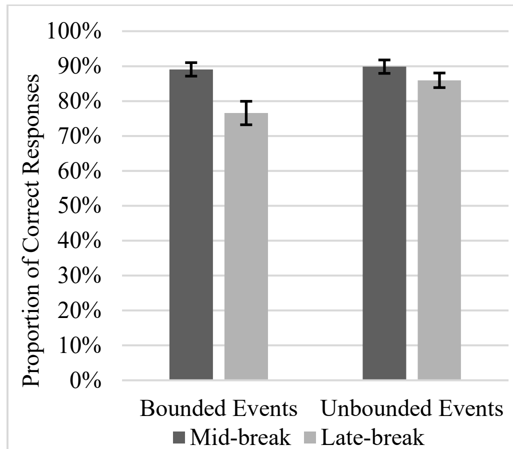

Figure 4: Proportion of correct responses in Experiment 2. Error bars represent ±SEM.

We further examined the response times for trials in which participants correctly identified the breaks in test items. The response times were analyzed using generalized linear mixed-effects models (GLMMs) with Event Type (Bounded vs. Unbounded) and Break Placement (Mid vs. Late) as fixed factors and crossed random intercepts for Subjects and Items. The models were fitted using the *glmer* function in R, and Gamma distribution was selected to provide a close approximation to the positively skewed distribution of response times (Lo & Andrews, 2015; R Core Team, 2013). As shown in Figure 5, participants spent more time on detecting a break in bounded events (M = 821 ms) compared to unbounded ones (M = 689 ms) ($\beta = -139.7$, $t = -3.23$, $p = .001$). Additionally, participants needed more time to detect breaks close to event endings (M = 796 ms) than at event midpoints (M = 714 ms) ($\beta = 62.05$, $t = 5.99$, $p < .001$). Importantly, a significant interaction between Event Type and Break Placement was found ($\beta = -36.78$, $t = -1.97$, $p = .049$). Participants watching bounded events had longer response times for late-breaks (M = 882 ms) compared to mid-breaks (M = 760 ms) ($\beta = 78.86$, $t = 5.27$, $p < .001$). The

[3] The accuracy of responses to both test items and fillers in Experiment 2 was significantly lower compared to Experiment 1 (Test items: $\beta = -0.62$, $z = -2.47$, $p = .013$; Fillers: $\beta = -1.23$, $z = -4.82$, $p < .001$).

difference in response times between mid-breaks (M = 669 ms) and late-breaks (M = 710 ms) became smaller when participants watched unbounded events ($\beta = 43.48$, $t = 3.10$, $p = .002$).

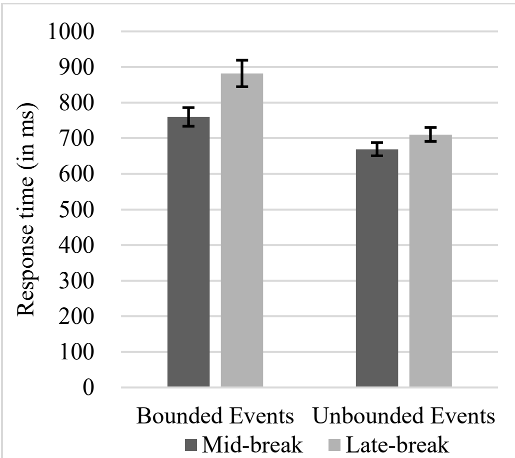

Figure 5: Response time (in ms) for participants to correctly identify a break in Experiment 2. Error bars represent ±SEM.

Discussion

Unlike Experiment 1, participants' accuracy was only affected by whether the break appeared in the middle or towards the end of the video, and the detection performance was overall lower. We hypothesize that these differences could have resulted from the change in the task: participants performed a more demanding dual task as they had to make a response during event perception (see also Papenmeier et al., 2019). Nevertheless, the patterns found in response times were reminiscent of the results from the previous experiments: participants took longer to detect disruptions close to event endings than at event middles, and this difference was greater in bounded than unbounded events. These results confirmed our hypothesis that boundedness affects online, spontaneous event perception.

General Discussion

Most studies on event cognition have typically used event segmentation measures but have paid less attention to the representational content of each event unit, or of classes of event units. Here we have used an innovative measure to probe sensitivity to event-general properties of events that was inspired by linguistic and philosophical treatments (e.g., Bach, 1986; Krifka, 1989, 1998; Vendler, 1957).

We hypothesized that, when people observe real-world events, they spontaneously construct coherent interpretations that incorporate the internal temporal contour of the events (i.e., boundedness) and use this information to process continuous visual input. In Experiments 1, we placed disruptions at different time points during bounded and unbounded naturalistic events and measured the accuracy of detecting these disruptions. The results showed that the placement of disruptions affected detection performance only for bounded events. In Experiment 2, we further measured the time it took to detect the disruptions as the event was unfolding. The results indicated that the disruption placement influenced response times to a greater extent in bounded events compared to the unbounded ones. These patterns confirmed our hypothesis: viewers spontaneously track the temporal texture of events as they make sense of dynamically unfolding event information.

Our results break new ground in studies of event cognition. First, they show that boundedness computations seem to be part and parcel of event comprehension rather than arising through the explicit, intentional extraction of commonalities among specific events. Second, the present findings reframe and contextualize a robust finding from prior studies on event segmentation, namely that event boundaries – especially event endpoints – are salient within the representation of an event (Hard et al., 2011; Huff et al., 2012; Newtson & Engquist, 1976; Pettijohn & Radvansky, 2016; Schwan & Garsofsky, 2004; Swallow et al., 2009; Zacks et al., 2007). Here we report that the relative salience of endpoints in event cognition is tied to the internal temporal texture of events and does not uniformly characterize event tokens. Last, the present data strongly suggest that boundedness should be integrated into existing models of event cognition. One possible path would be to recruit and enhance the mechanisms outlined in EST (Swallow et al., 2009; Zacks et al., 2007). Our results indicate that viewers are sensitive to accumulating change *within* the boundaries of an event, even when no event breakpoint is detected. Furthermore, depending on how predictable this change is, viewers construct different event types. During unbounded events, observers can easily predict what comes next based on what is happening in the moment, and treat temporal slices of the event similarly. By contrast, during bounded events, different temporal slices represent different stages of development.

Our present work leaves several questions open for future research. Even though the present stimuli were created to be unambiguously bounded or unbounded, in both language and cognition the same experience can often be construed from both a bounded and an unbounded perspective (compare *playing music* and *playing a musical piece*; Wagner & Carey, 2003). Furthermore, considerations of boundedness may interact with the agent's preferences, goals and other aspects of the context (Depraetere, 2007; Filip, 2001; Kennedy & Levin, 2008; Mathis & Papafragou, 2020; Zacks & Swallow, 2007). Further research needs to address *how* the viewer's mind extracts boundedness categories from streams of sensory information, and *how* this process affects information-processing at distinct temporal points along the development of the event.

Acknowledgments

This material is based upon work supported by the Beijing Institute of Technology Research Fund Program for Young Scholars (Y.J.) and National Science Foundation Grant BCS-2041171 (A.P.).

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy, 9*(1)*,* 5–16.

Barr, D. (2008). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Carruthers, P. (2017). Mindreading in adults: evaluating two-systems views. *Synthese, 194*(3)*,* 673–688.

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

Depraetere, I. (2007). (A)telicity and intentionality. *Linguistics, 45*(2), 243–269.

Filip, H. (2001). Nominal and verbal semantic structure: analogies and interactions. *Language Sciences, 23*(4), 453–501.

Filip, H. (2012). Lexical aspect. In R. I. Binnich (Ed.), *The Oxford handbook of tense and aspect* (pp. 721-751). Oxford University Press.

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586–604.

Huff, M., & Papenmeier, F. (2017). Event perception: From event boundaries to ongoing events. *Journal of Applied Research in Memory and Cognition, 6*(2), 129–132.

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

Ji, Y., & Papafragou, A. (2020). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition, 197,* 104197.

Kennedy, C., & Levin, B. (2008). Measure of change: The adjectival core of degree achievements. In L. McNally, & C. Kennedy (Eds.), *Adjectives and adverbs: Syntax, semantics and discourse*. Oxford University Press.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expression, Groningen-Amsterdam studies in semantics* (Volume 11). Foris Publications.

Krifka, M. (1998). The origins of telicity. In S. Rothstein (ed.), *Events and grammar*. Kluwer.

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology¸6*, 1-16.

Malaia, E. Renaweera, R., Wilbur, R., & Talavage, T. (2012). Event segmentation in a visual language: Neural bases of processing American Sign Language predicates. *Neuroimage, 59*(4)*,* 4094–4101.

Mathis, A., & Papafragou, A. (2020, March 19). Agents' goals affect placement of event endpoints [Poster session]. The 33rd Annual CUNY Conference on Human Sentence Processing, UMass-Amherst, Amherst, MA, United States.

Mittwoch, A. (2013). On the criteria for distinguishing accomplishments from activities, and two types of aspectual misfits. In B. Arsenijević, B. Gehrke & R. Marín (Eds.), *Studies in the composition and decomposition of event predicates*. Springer.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*(5)*,* 436–450.

O'Grady, C., Scott-Phillips, T., Lavelle, S., & Smith, K. (2020). Perspective-taking is spontaneous but not automatic. *Quarterly Journal of Experimental Psychology*, *73*(10), 1605–1628.

Papenmeier, F., Brockhoff, A., & Huff, M. (2019). Filling the gap despite full attention: The role of fast backward inferences for event completion. *Cognitive Research: Principles and Implications, 4,* 1–17.

Parsons, T. (1990). *Events in the semantics of English*: *A study in subatomic semantics*. MIT Press.

Pettijohn, K. A., & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation. *Memory & Cognition, 44*(7), 1064-1075.

Radvansky, G. A., & Zacks, J. (2014). *Event cognition*. Oxford University Press.

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Available online http://www.R-project.org

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18*(1), 37–55.

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M. & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences of the United States of America, 112*(19), 5968–5973.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distorts memory in a matter of seconds. *Cognition, 121*(3)*,* 409–415.

Swallow, K., Zacks, J., & Abrams, R. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*(2), 236–257.

Tenny, C. (1987). *Grammaticalizing aspect and affectedness* [Unpublished doctoral dissertation]. MIT.

van Hout, A. (2016). Lexical and grammatical aspect. In J. Lidz, W. Synder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics*. Oxford University Press.

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

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition, 90*(2)*,* 163–191.

Wehry, J., Hafri, A., & Trueswell, J. (2019). The end's in plain sight: Implicit association of visual and conceptual boundedness. In A.K. Goel, C.M. Seifert, & C. Freksa

(Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Wellwood, A., Hespos, S. J., & Rips, L. (2018). The object : substance :: event : process analogy. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy* (Volume 2). Oxford University Press.

Zacks, J., Speer, N., Swallow, K., Braver, T., & Reynolds, J. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133* (2), 273–293.

Zacks, J., & Swallow, M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16,* 80–84.

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD83