Boundedness in Event Cognition:

Viewers Spontaneously Represent the Temporal Texture of Events

Yue Ji[a] and Anna Papafragou[b]

a. Department of English, School of Foreign Languages, Beijing Institute of Technology. No. 5 South Street, Zhongguancun, Haidian District, Beijing, China, 100081. E-mail: jiyue@bit.edu.cn

b. Department of Linguistics, University of Pennsylvania. 3401-C Walnut St., Philadelphia, PA, USA, 19104. E-mail: anna4@sas.upenn.edu

Author Note

Yue Ji https://orcid.org/0000-0002-9978-4006

Anna Papafragou https://orcid.org/0000-0001-5435-1058

Correspondence concerning this article should be addressed to Yue Ji, Department of English, School of Foreign Languages, Beijing Institute of Technology, Beijing, 100081. E-mail: jiyue@bit.edu.cn

Abstract

A long philosophical and linguistic literature on events going back to Aristotle distinguishes between events that are internally structured in terms of distinct temporal stages leading to culmination (*bounded* events; e.g., a girl folded up a handkerchief) and events that are internally unstructured and lack an inherent endpoint (*unbounded* events; e.g., a girl waved a handkerchief). Here we show that event cognition spontaneously computes this foundational dimension of the temporal texture of events. People watched videos of either bounded or unbounded events that included a visual interruption lasting either .13s (Experiment 1) or .03s (Experiments 2 and 3). The interruption was 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 and 2) or respond as soon as they detected an interruption while watching each video (Experiment 3). When people responded after the video, they were more likely to ignore interruptions placed close to event endpoints compared to event midpoints (Experiment 1); similarly, when they responded during the video, they reacted more slowly to endpoint compared to midpoint interruptions (Experiment 3). Crucially, across the three experiments, there was an interaction between event type and interruption timing: the endpoint-midpoint difference depended on whether participants were watching an event that was bounded or unbounded. These results suggest that, as people perceive dynamic events, they spontaneously track boundedness, or the temporal texture of events. This finding has implications for current models of event cognition and the language-cognition interface.

Keywords: events, event segmentation, boundedness, aspect

Boundedness in Event Cognition:

Viewers Spontaneously Represent the Temporal Texture of Events

Introduction

Our experience of the world involves dynamic, continuous streams of visual input but human cognition spontaneously and rapidly organizes this input into coherent and discrete event units. According to a prominent account (Event Segmentation Theory, or EST; Zacks et al., 2007), the process of segmenting events from continuous streams of actions 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 in the input stream. In the framework of EST, 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. The changes that lead to prediction errors – and hence to the placement of event boundaries - can be perceptual, such as a change of location for event participants (e.g., a student coming home from school; Magliano et al., 2001; Newtson et al., 1977; Zacks et al., 2006), or conceptual, such as the achievement of a goal (e.g., a student coming up with a solution to a math problem; Zacks & Swallow, 2007).

A key finding from the literature on event cognition is that event boundaries are influential for event processing. For instance, deletions of event boundaries are more noticeable compared to deletions of non-boundary moments (Newston & Engquist, 1976); furthermore, visual stimuli that include only event boundaries are understood and recalled better than stimuli that include only event middles (ibid.; see also 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 from the end of one event to the beginning of the next 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; Hard et al., 2019; 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, in other words, the intuitive notion of "what happens" between the time an event begins and ends. 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. Despite being foundational for the nature of events, and a cornerstone of the logical and linguistic analysis of events, the temporal texture has largely been absent from current cognitive event frameworks.

Boundedness in language and cognition

According to a long linguistic and philosophical study of events that goes back to Aristotle (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, p. 146) – 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, 2020a), 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. Furthermore, when asked to indicate what kind of event was marked by a red frame, they were likely to mention the structure and organization (or lack thereof) of the events, thereby showing sensitivity to an essential dimension of the bounded-unbounded distinction (cf. also Ji & Papafragou, 2020b). Other studies have offered evidence that boundedness cross-cuts linguistic and visual stimuli (e.g., Malaia et al., 2012; Strickland et al., 2015; Wagner & Carey, 2003; Wehry et al., 2019; Wellwood et al., 2018).

These new findings point to a rich theoretical model of event structure and event boundaries that can naturally bridge insights from psychological and linguistic perspectives on events. Going beyond the influential view that an event is “a segment of time at a given location that is conceived by an observer to have a beginning and an end” (Zacks & Tversky, 2001), the notion of boundedness captures the fundamental intuition that events are internally organized in different ways, and as such can come to an end differently. For bounded events (e.g., fix the car), the endpoint indicates culmination that often coincides with a moment of maximal change in the state of the object affected by the event (here, the car). For unbounded events (e.g., drive the car), the endpoint is a simple point of transition to a different event or state of affairs (e.g., park the car) and does not coincide with a change in the object involved in the event. Both types of endpoints would be classified as event boundaries in traditional accounts of event segmentation but they represent something very different in terms of what happens in each case.

The nature of boundedness computations

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). Recall that, 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 homogeneous (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 since they are equally predictable. By contrast, during non-homogeneous (bounded) 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

[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.).

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, 2020a). 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.

Such an outcome would be theoretically important: if boundedness is computed spontaneously during event understanding, it could offer a powerful way of organizing incoming event information, readily connect to the way events are encoded in language, and potentially have further effects on event cognition. For instance, recall that, according to a large body of work in cognitive psychology, event boundaries – and especially, event endpoints - are privileged in both event comprehension and memory compared to other temporal slices of events (Hard et al., 2011; Hard et al., 2019; Huff et al., 2012; Newtson & Engquist, 1976; Pettijohn & Radvansky, 2016; Schwan & Garsofsky, 2004; Swallow et al., 2009; Zacks et al., 2007). However, past studies on event segmentation have not differentiated between event types, and have typically focused on just bounded examples without clearly motivating this choice. If (un)boundedness is spontaneously computed by event cognition, event endpoints should be privileged compared to other time points within an event only (or particularly) for bounded events: the transition from one stage of such events to the next involves significant internal change and thus requires more processing resources. For unbounded events that lack internal transitions, the difference between endpoints and other temporal time points should be smaller or non-existent (for evidence from an explicit category learning task, see Ji and Papafragou, 2020b).

The present study

In the present study, we tested the hypothesis that event cognition spontaneously tracks the temporal texture of bounded and unbounded events using a novel task. Our broader goal was to contribute to a unified theory of event representation that bridges insights from psychological and linguistic perspectives. We introduced very brief disruptions at different time points within videos of bounded vs. unbounded events during which the visual stimulus became blurry. Prior work has found that the recall and recognition of such disruptions is very poor (e.g., Levin & Varakin, 2004). Here we adopted an explicit detection task where the observers were told to find disruptions from the beginning (see also Huff et al., 2012). Observers had to respond either after watching a video (Experiments 1 and 2) or as soon as they detected the interruption as they watched the video (Experiment 3). Since disruptions were irrelevant to event content, 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, as we expect, boundedness is computed as part of event apprehension, even when not required by the observer's immediate task, 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. [2]

Data availability

The data, analysis codes and stimuli for the present study are available through the Open Science Framework (https://osf.io/4csrq/, "Temporal texture of events" project).

Experiment 1

Method

Participants

Sixty-four adults (age range: 18-23) participated in the experiment. All participants were undergraduates at a major university on the East Coast of the US. Participants signed an informed consent form approved by the institutional review board of the university and received course credit for participation. Data from 3 additional adults were collected but excluded because they kept giving *Break* responses throughout the experiment. The sample size of Experiment 1 was decided based on the sample size in similar studies on event perception and memory (Huff et al., 2012; Papenmeier et al., 2019) and a prior related study (Experiment 1 in Ji & Papafragou, 2020a).

Stimuli

We used the same 20 pairs of videos as Ji and Papafragou (2020a). A blank screen was

[2] Recall that in earlier work by Newtson and Engquist (1976), omissions of event boundaries were more noticeable compared to omissions of non-boundary moments for what we would classify as bounded events. In that study, participants had to examine "continuous behavior" and detect possible missing action or action parts. By contrast, our study simply asked people to detect "breaks", a task irrelevant to the content of the video: for bounded events, we expected this task to be harder at endpoints compared to midpoints. Therefore, the two patterns could be interpreted as complementary to each other: our predictions were based on the idea that, at event boundaries, observers are sensitive to changes relevant to the event content and tend to miss irrelevant changes.

displayed for 0.8s at the beginning and the end of each video. Paired videos showed a bounded and an unbounded event; within each pair, the videos were purposefully matched in duration. Stimuli duration ranged from 4.4 to12.0s (M = 7.8s, SD = 2.4; 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 Filip, 2004; Tenny, 1987). For half of the videos, paired bounded and unbounded events 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., fold up a handkerchief) while its unbounded counterpart did not involve such a change (e.g., wave a handkerchief). 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).

Table 1

Event stimuli in Experiment 1.

Phase	No.	Bounded Events	Unbounded Events	Duration	Boundedness Source
Practice	1	close a fan	use a fan for oneself	4.40s	Nature of Action
	2	crack an egg	beat an egg	6.00s	
	3	cut a ribbon in half	cut ribbon from a roll	6.40s	Nature of Affected Object
	4	stick a sticker	stick stickers	4.67s	
Testing	5	fold up a handkerchief	wave a handkerchief	8.00s	Nature of Action
	6	put up one's hair	scratch one's hair	8.00s	
	7	stack a deck of cards	shuffle a deck of cards	6.33s	
	8	group pawns based on color	mix pawns of two colors	7.50s	
	9	dress a teddy bear	pat a teddy bear	12.00s	
	10	roll up a towel	twist a towel	7.50s	
	11	fill a glass with milk	shake a bottle of milk	8.27s	
	12	scoop up yogurt	stir yogurt	5.33s	
	13	draw a balloon	draw circles	8.00s	Nature of Affected Object
	14	tie a knot	tie knots	7.00s	
	15	eat a pretzel	eat cheerios	12.00s	
	16	flip a postcard	flip pages	4.67s	
	17	peel a banana	crack peanuts	11.13s	
	18	blow a balloon	blow bubbles	9.00s	
	19	tear a paper towel	tear paper towels	8.00s	
	20	paint a star	paint stuff	11.33s	

Four types of norming studies were conducted on these stimuli. First, a separate group of 18 English native speakers described the video clips (Ji & Papafragou, 2020a). Their descriptions underwent linguistic tests for boundedness (e.g., bounded descriptions can be modified by delimited temporal phrases such as *in an hour* while unbounded descriptions go along with durative temporal phrases such as *for an hour*; see Dowty, 1979; Vendler, 1957). The results showed that the videos successfully aligned with the linguistic boundedness distinction: stimuli of bounded events elicited bounded descriptions that included change-of-state predicates (e.g., fold up a handkerchief) 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., wave a handkerchief) or unquantified noun phrases (bare plurals or mass nouns: e.g., blow bubbles) 92.8% of the time. There was no significant difference between the two event types in terms of whether they elicited the corresponding aspectual distinctions in the production task (t (17) = 1.84, p = .083). Second, a new group of 40 participants provided judgments about the temporal structure of the stimuli. Videos of bounded events were considered as "something with a beginning, midpoint and specific endpoint" 87% of the time while videos of unbounded events were considered as such only 21.5% of the time (a significant difference, $t(39) = 20.33$, $p < .0001$). These two norming studies confirmed that observers talked about and explicitly judged our stimuli as either bounded or unbounded events as expected. Third, a new group of 20 participants rated how intentional the action in each video looked on a scale from 1 (totally unintentional) to 7 (intentional) (Ji & Papafragou, 2020a). The degree of intentionality did not differ between bounded (M = 5.67) and unbounded (M = 5.62) events (t (19) = 1.34, p = .195). Finally, another group of 40 participants rated the speed of the girl's action in the middle or at the end of the video on a scale from 1 (very slow) to 7 (very fast). The ratings did not differ between bounded (M = 4.23) and unbounded (M = 4.27) events (F (1, 38) = 0.041, p > .250), or between the middle (M = 4.26) and the end (M = 4.25) of the video (F (1, 38) = .005, p > .250). No significant interaction between the two factors was detected (F (1, 38) = .740, p > .250).

The videos were edited in Corel VideoStudio X9 to introduce a “break” of 0.13s (i.e., 4 editing frames, 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 4 frames that showed the temporal midpoint of the event (e.g., in the example of folding up a handkerchief consisting of 240 frames, the break replaced the 119^{th}, 120^{th}, 121^{st}, and 122^{nd} frames). In the end-break version, the break blocked the last four frames of the event. Since the videos showed the actor manipulate object(s) with her hand (e.g., folding up a handkerchief), or other body part (e.g., blowing a balloon), the end-break blocked the moment when the actor’s body part got separated from the object(s). Edited videos were used as test items, and their original (unedited) versions were used as fillers.

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 an end-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 end-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 an end-break) and 8 fillers. Similar to the practice phase, 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. end) 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 tested in groups of four in the lab. An experimenter gave them instructions and showed the videos on a projector. Participants were requested to watch each video carefully and decide whether they saw a break in the video. Responses were given after the end of each video by circling either “Break”, or “No break” on an answer sheet. Participants were 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 any participant in the group was wrong, the video was played a second time. In the testing phase, no feedback was given.

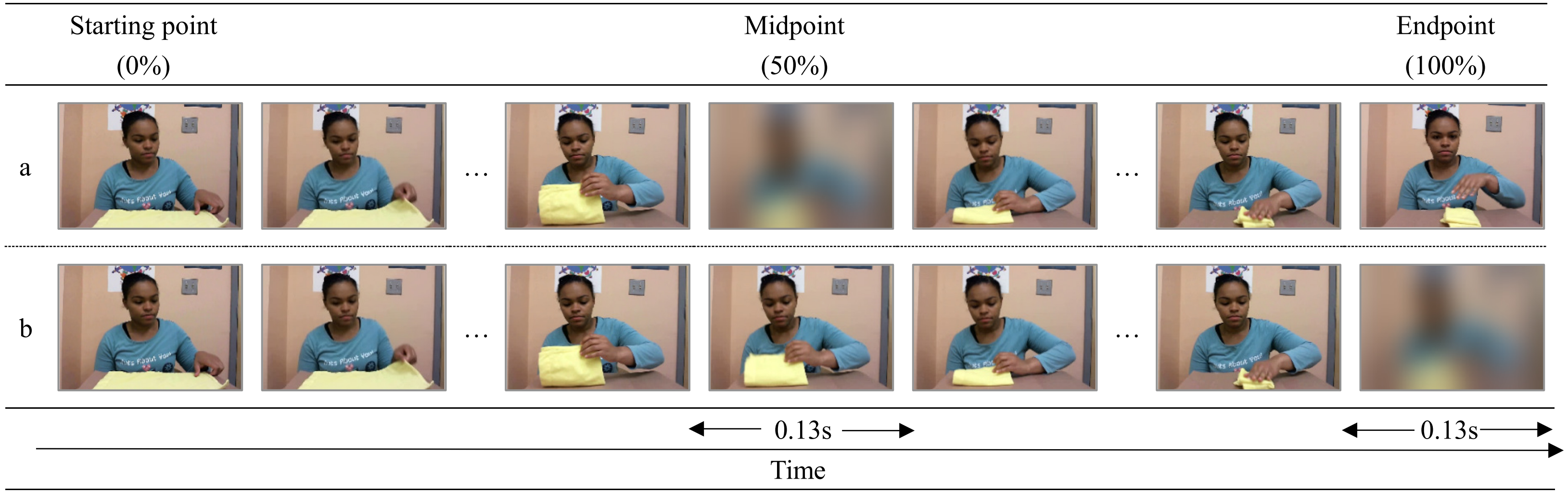

Figure 1. Examples of two versions of a bounded event (fold up a handkerchief) in Experiment 1: (a) mid-break (b) end-break.

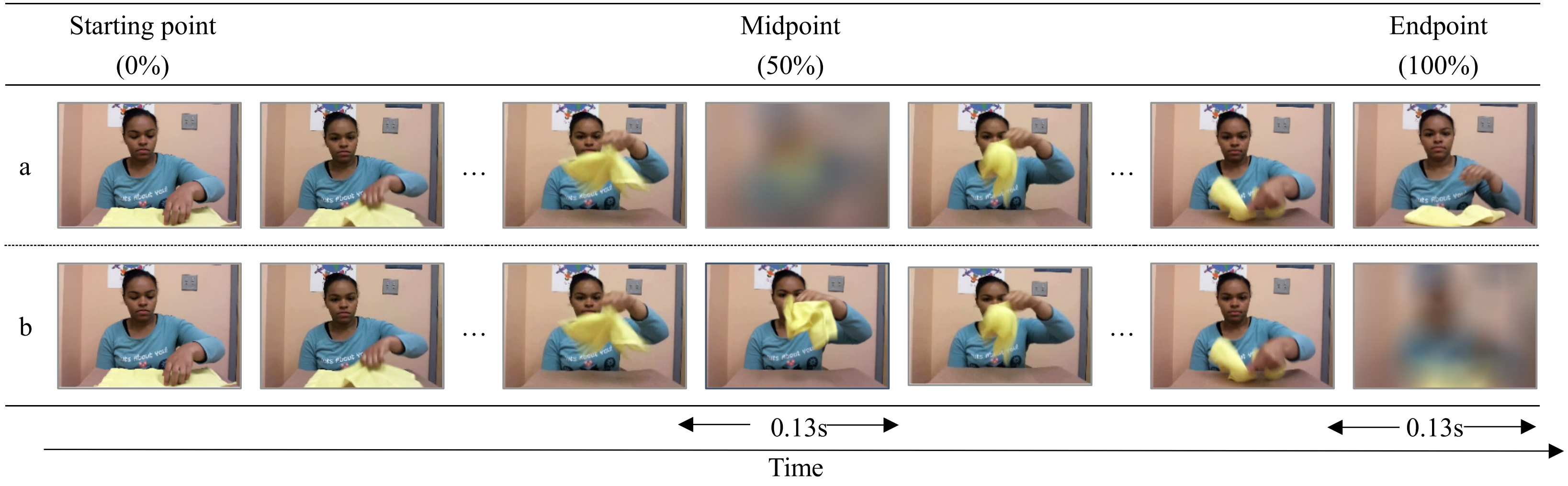

Figure 2. Examples of two versions of an unbounded event (wave a handkerchief) in Experiment 1: (a) mid-break (b) end-break

Results

"Break" responses to test items and "No break" responses to fillers were coded as correct. We analyzed the binary accuracy data using mixed-effects modeling. Random intercepts were provided for each Subject and each Item (Baayen, Davidson, & Bates, 2008; Barr, 2008). All models were fitted using the *glmer* function of the *lme4* package in R (R Core Team, 2013). Two-level categorical predictors were coded using centered contrasts (-0.5, 0.5). Overall, the accuracy of responses to fillers did not differ significantly between the Bounded (M = 91.8%, SE = 0.020) and Unbounded condition (M = 95.7%, SE = 0.011) ($\beta = 0.75$, $z = 1.33$, $p = .183$). Turning to test items (see Figure 3), we examined the fixed effects of Event Type (Bounded vs. Unbounded), Break Placement (Mid vs. End) and their interaction.[3] In addition, non-theoretically-driven predictors, including List, Gender, Boundedness Source (Action vs. Affected Object), and any interaction between Boundedness Source and other predictors were added incrementally to the model. These predictors or interactions did not reliably improve model fit (assessed by chi-square tests on the log-likelihood values of competing models with three indices, AIC, BIC and logLik), and were excluded from further analysis. The same strategy of model selection was applied to the analyses in the following experiments. In the final model of this experiment, the random intercept of Item was not included because the estimated random effect for Item was close to zero. The average performance on each item can be found in the

[3] Based on the guidelines in Barr et al. (2013, p. 275) and follow-up suggestions in Barr (2013, p. 1), we included a random slope for the within-subjects factor Break Placement when building models incrementally. However, models that included the random slope either did not converge, or failed to improve the model fit. Therefore, we kept only random intercepts in our final models. The same treatment of random slopes can be found in similar work on the interface between event language and event cognition (e.g., Kuhn et al., 2021; Lee & Kaiser, 2021).

Appendix.

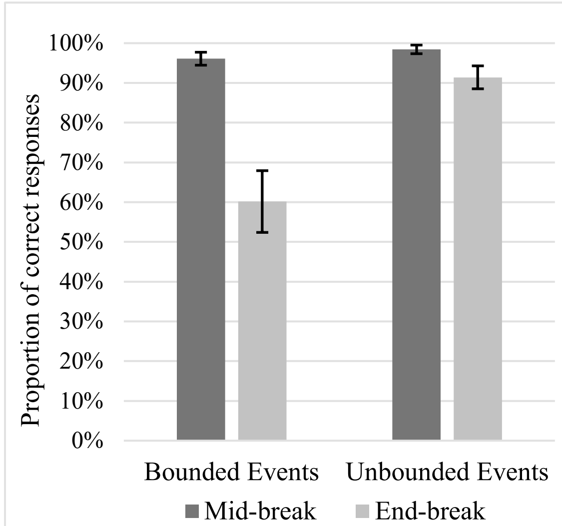

Figure 3. Proportion of correct responses in Experiment 1. Error bars represent ±SEM.

As shown in Table 2, both Event Type and Break Placement were significant. Participants were more successful in detecting the break in unbounded events (M = 94.9%, SE = 0.015) than in bounded events (M = 78.1%, SE = 0.042) (β =1.70, z = 2.01, p = .045). Furthermore, breaks at midpoints (M = 97.3%, SE = 0.010) were better identified than breaks at endpoints (M = 75.8%, SE = 0.046) (β = -3.22, z = -5.79, p < .001). Importantly, there was a significant interaction between Event Type and Break Placement (β = 2.45, z = 2.26, p =.024). When participants watched videos of bounded events, they were better at detecting mid-breaks (M = 96.1%, SE = 0.004) than end-breaks (M = 60.2%, SE = 0.112) (β = -4.80, z = -6.03, p <.001). This pattern was also found in unbounded events (β = -1.82, z = -2.45, p = .014), but the difference between mid-breaks (M = 98.4%, SE = 0.003) and end-breaks (M = 91.4%, SE = 0.015) was smaller.

Table 2

Fixed effect estimates for multi-level model of accuracy in break detection in Experiment 1.

Effect	Estimate	SE	z value
(intercept)	3.99	0.58	6.90***
Event Type (Bounded vs. Unbounded)	1.70	0.85	2.01*
Break Placement (Mid vs. End)	-3.22	0.56	-5.79***
Event Type*Break Placement	2.45	1.08	2.26*

Note. Formula in R: Acc ~ 1 + (1|Subject) + Event Type + Break Placement + Event Type : Break Placement
* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

Results from Experiment 1 show that viewers tended to miss visual breaks at the endpoint compared to the midpoint of dynamically unfolding events, in accordance with previous work pointing to the importance of event endpoints for event comprehension (Newston & Engquist, 1976; Schwan & Garsofsky, 2004; Swallow et al., 2009; Zacks et al., 2007; Hard et al., 2011; Hard et al., 2019; Pettijohn & Radvansky, 2016; Huff et al., 2012). Importantly, however, this effect was more potent for bounded (structured, non-homogeneous) compared to unbounded (non-structured, homogeneous) events. Furthermore, differential sensitivity to the placement of visual interruptions emerged despite the fact that the participants' task was simply to attend to and detect the visual breaks and did not require processing the specifics of the event contents. The computation of boundedness was at least partly event-general, since it did not differ depending on whether the event was considered bounded or not on the basis of the action or the affected object.

Together, these findings suggest that event cognition spontaneously computes the abstract internal temporal texture (or boundedness) of events, and this computation affects the way incoming event streams are processed. This conclusion is consistent with but goes considerably beyond prior evidence on the contribution of boundedness to event cognition (cf. also Ji & Papafragou, 2020a, 2020b; Strickland et al., 2015, a.o.).

Could the differences in disruption detection be due to some aspect of the stimuli other than boundedness? We believe that this is unlikely, since prior norming ensured that the classes of bounded and unbounded events contained events that were equally intentional. Furthermore, prior norming showed that the action in both classes of events was judged as equally fast at critical time points (i.e., the middle or the end of the video). Other potential visual correlates of temporal structure such as repetition did not uniquely characterize one or the other boundedness class in our stimuli (e.g., one third of bounded events involved a repetitive action; cf. Ji & Papafragou, 2020a).

Experiment 2

If the asymmetries observed in Experiment 1 characterize how viewers track ongoing event developments, such asymmetries should surface even before the actual endpoint of an event is reached. In Experiment 2, we compared the detection of visual breaks placed at midpoints vs. time points close to the endpoints. For bounded events that have distinct sub-stages leading to a moment of culmination, break detection at event middles should be better compared to break detection close to event endings, as observers' attention would be drawn towards what was achieved at the later timepoint. By contrast, for unbounded events that have a largely undifferentiated temporal structure, the difference in break detection should diminish or disappear. We tested this expectation in a version of Experiment 1 in which end-breaks were replaced by late-breaks (beginning at 80% of the video but before the moment the action stopped or culminated). To make the task more challenging, especially since the late-breaks were no longer at the very end of the event stream, we shortened mid- and late-breaks to just one editing frame (0.03s).

Method

Participants

Sixty-four adults (age range: 18-22) were recruited from the same population as Experiment 1 and received course credit for participating. Data from two additional adults were collected but excluded because they did not understand what a "break" was even after the practice phase. The sample size of Experiment 2 was decided by running a power analysis of Experiment 1. With 64 participants, the observed power of the predictor of interest – the interaction between Event Type and Break Placement – was 0.87.

Stimuli

Video stimuli were the same as Experiment 1 with two exceptions. First, the duration of the breaks was shortened to 0.03s (or 1 editing frame). There was little change in the action after this extremely brief break (see Figures 4 and 5). Second, mid-breaks were centered around the point corresponding to 50% of the video (e.g., in the example of folding up a handkerchief consisting of 240 frames, the break replaced the 121st frame). Late-breaks began at the point that corresponded to 80% of the video (e.g., for the same event of folding up a handkerchief, the break replaced the 193rd frame).

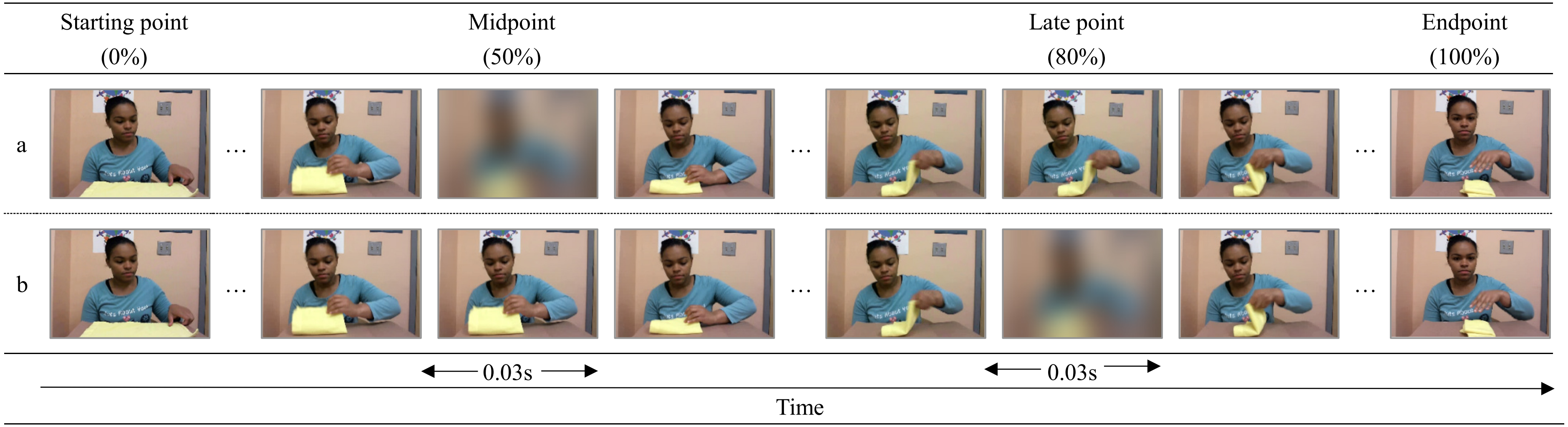

Figure 4. Examples of two versions of a bounded event (fold up a handkerchief) in Experiment 2: (a) mid-break (b) late-break.

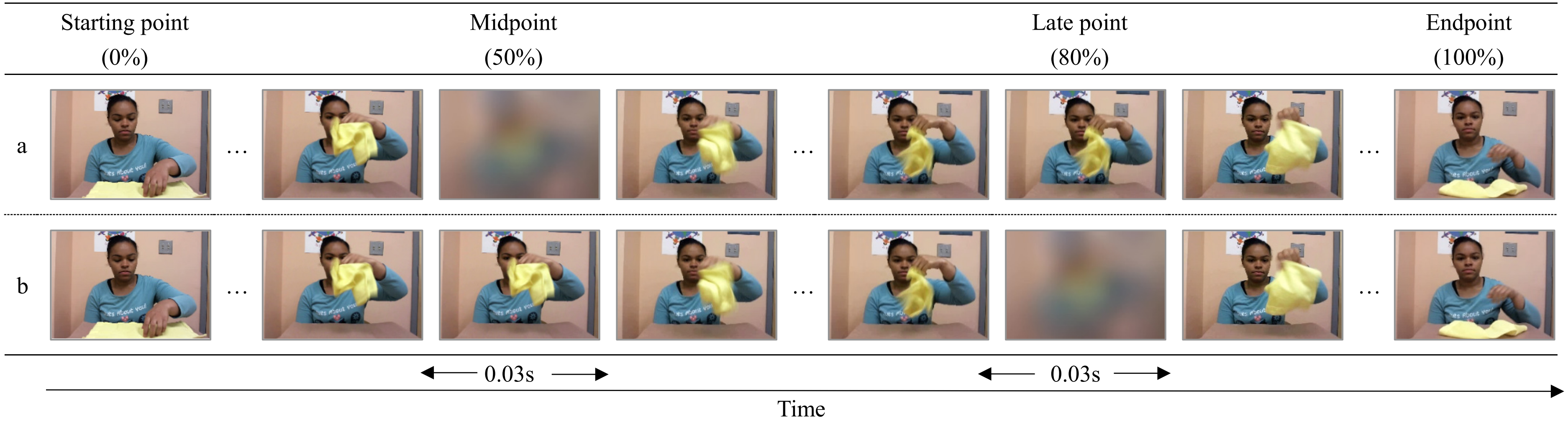

Figure 5. Examples of two versions of an unbounded event (wave a handkerchief) in Experiment 2: (a) mid-break (b) late-break.

Procedure

The procedure was identical to Experiment 1.

Results

The same coding and analytic strategy was used as in Experiment 1. Performance on filler items did not significantly differ between event types (M = 93.8%, SE = 0.020 for Bounded vs. M = 92.2%, SE = 0.016 for the Unbounded events, $\beta = 0.25$, $z = 0.56$, $p > .250$). Turning to test items, the binary accuracy of detecting breaks was analyzed using a mixed logit model with the fixed effects of Event Type (Bounded vs. Unbounded), Break Placement (Mid vs. Late) and their interaction. Random intercepts were provided for each Subject and each Item. The analysis (Figure 6 and Table 3) showed that the difference between Bounded (M = 87.5%, SE = 0.025) and Unbounded event types (M = 94.5%, SE = 0.016) was not significant ($p = .072$). The difference in break detection between midpoints (M = 94.5%, SE = 0.016) and late points (M = 87.5%, SE = 0.025) did not reach significant either ($p = .061$). However, there was a significant interaction between Event Type and Break Placement ($\beta = 1.99$, $z = 2.70$, $p = .007$). Participants watching videos of bounded events were better at detecting mid-breaks (M = 95.3%, SE = 0.019) than late-breaks (M = 79.7%, SE = 0.042) ($\beta = -1.74$, $z = -3.50$, $p < .001$). By contrast, participants watching videos of unbounded events did not differ in their detection of mid-breaks (M = 93.8%, SE = 0.023) and late-breaks (M = 95.3%, SE = 0.019) ($\beta = 0.31$, $z = 0.55$, $p > .250$).

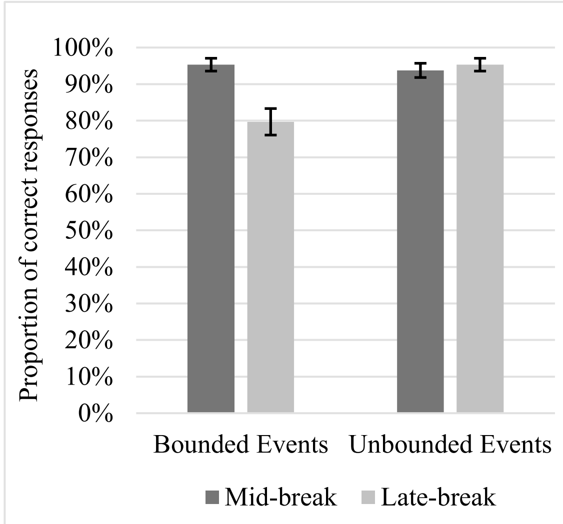

Figure 6. Proportion of correct responses in Experiment 2. Error bars represent ±SEM.

Table 3

Fixed effect estimates for multi-level model of accuracy in break detection in Experiment 2.

Effect	Estimate	SE	z value
(intercept)	2.62	0.26	10.03***
Event Type (Bounded vs. Unbounded)	0.69	0.38	1.80
Break Placement (Mid vs. Late)	-0.69	0.37	-1.87
Event Type*Break Placement	1.99	0.74	2.70**

Note. Formula in R: Acc ~ 1 + (1|Subject) + (1|Item) + Event Type + Break Placement + Event Type : Break Placement
* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The main result of Experiment 2 is that 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. Together with the results from Experiment 1, these data support the conclusion that observers track the temporal texture of events as part of their event understanding.

Experiment 3

In Experiments 1 and 2, 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 3, 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 pressed the spacebar (indicating that a break was detected) 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. The sample size of Experiment 3 was justified by a power analysis using data from Experiment 2. With 64 participants, the observed power of the predictor of interest – the interaction between Event Type and Break Placement – was 0.92.

Stimuli

Video stimuli were the same as Experiment 2.

Procedure

Experiment 3 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 an interruption, or 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, the whole video was played despite the fact that participants might have pressed the Spacebar or N before the end of the video. Both the response type and response time were recorded. As long as participants made a response within 5 seconds after a video ended, a "Done" button would appear under the video and participants clicked it to advance to the next trial. If no response was given within the 5 seconds after a video ended, the program automatically moved on to the next trial. As in Experiments 1 and 2, 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

Accuracy of responses

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 = 8, 0.8% 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%, SE = 0.034 for Bounded vs. M = 78.9%, SE = 0.036 for the Unbounded events, $\beta = -0.09$, $z = -0.31$, $p > .250$). For test items, the same coding and analytic strategy was used as in Experiments 1 and 2. As shown in Figure 7 and Table 4, there was a significant effect of Break Placement, such that participants were better at detecting breaks at midpoints (M = 89.5%, SE = 0.022) than breaks close to event endpoints (M = 81.3%, SE = 0.033) ($\beta = -0.72$, $z = -2.60$, $p = .009$). Unlike the previous experiments, the difference between Bounded (M = 82.8%, SE = 0.033) and Unbounded (M = 87.9%, SE = 0.026) event types was not significant ($\beta = 0.37$, $z = 1.04$, $p > .250$), nor was there a significant interaction between Event Type and Break Placement ($\beta = 0.63$, $z = 1.14$, $p > .250$).[4]

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

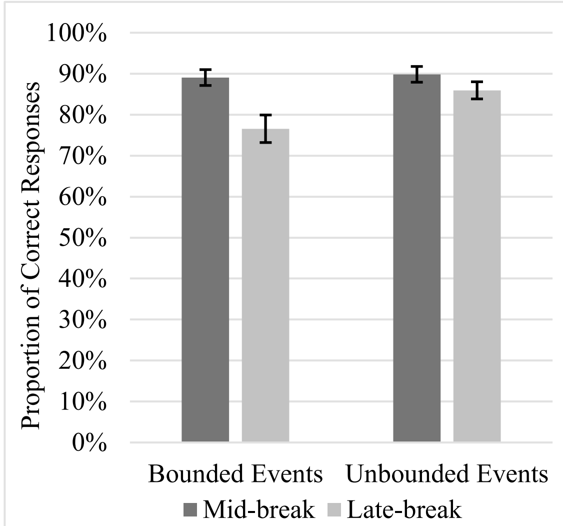

Figure 7. Proportion of correct responses in Experiment 3. Error bars represent ±SEM.

Table 4

Fixed effect estimates for multi-level model of accuracy in break detection in Experiment 3.

Effect	Estimate	SE	z value
(intercept)	2.14	0.25	8.47***
Event Type (Bounded vs. Unbounded)	0.37	0.36	1.04
Break Placement (Mid vs. Late)	-0.72	0.28	-2.60**
Event Type*Placement	0.63	0.55	1.14

Note. Formula in R: Acc ~ 1 + (1|Subject) + (1|Item) + Event Type + Placement + Event Type : Placement

* $p < .05$, ** $p < .01$, *** $p < .001$

Response times

We further examined the response times for trials in which participants correctly identified the breaks in test items. In 77 trials (17.3% of the total correct test items), participants pressed the spacebar more than once. Most responses that followed the first response occurred after the end of the video. In these cases, we only included the first response time in our analysis.

In 8 trials, however, participants responded at least twice before the end of the video, and these were excluded from further analysis. After exclusions, there were 437 response times that were entered in the following analysis.

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. Gamma distribution with the identity link function 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 8 and Table 5, participants spent more time on detecting a break in bounded events (M = 821 ms, SE = 38.6) compared to unbounded ones (M = 689 ms, SE = 33.6) ($\beta = -139.7$, $t = -3.90$, $p < .001$). Additionally, participants needed more time to detect breaks close to event endings (M = 796 ms, SE = 31.1) than at event midpoints (M = 714 ms, SE = 23.0) ($\beta = 62.05$, $t = 5.90$, $p < .001$). Importantly, a significant interaction between Event Type and Break Placement was found ($\beta = -36.78$, $t = -2.01$, $p = .045$). Participants watching bounded events had longer response times for late-breaks (M = 882 ms, SE = 35.3) compared to mid-breaks (M = 760 ms, SE = 25.9) ($\beta = 78.86$, $t = 5.42$, $p < .001$). The difference in response times between mid-breaks (M = 669 ms, SE = 18.7) and late-breaks (M = 710 ms, SE = 20.1) became smaller in participants who watched unbounded events ($\beta = 43.48$, $t = 3.36$, $p = .001$).

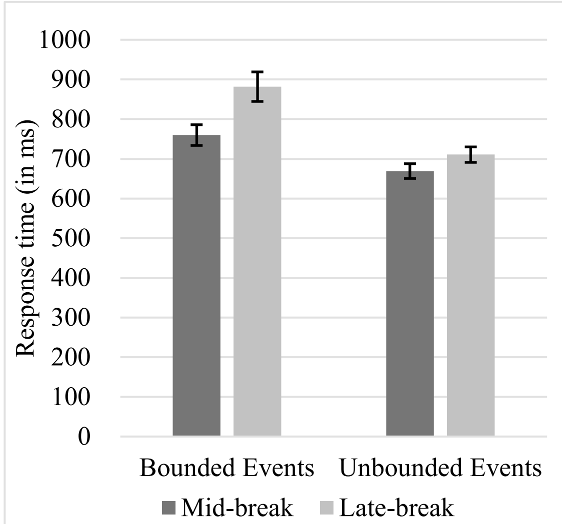

Figure 8. Response time (in ms) for participants to correctly identify a break in Experiment 3. Error bars represent ±SEM.

Table 5

Fixed effect estimates for multi-level model of response times (in ms) in Experiment 3.

Effect	Estimate	SE	*t* value
(intercept)	796.96	29.12	27.18 ***
Event Type (Bounded vs. Unbounded)	-139.66	35.81	-3.90***
Break Placement (Mid vs. Late)	62.05	10.52	5.90***
Event Type*Placement	-36.78	18.32	-2.01*

Note. Formula in R: RT ~ 1 + (1|Subject) + (1|Item) + Event Type + Placement + Event Type : Placement
* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

Unlike Experiments 1 and 2, there was no interaction between event type (bounded vs. unbounded) and break placement (mid- vs, late-break) on participants' successful break detection; participants' accuracy was only affected by whether the break appeared in the middle or towards the end of the video. In addition, the detection performance was overall lower compared to Experiment 2 (see fn.5). 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; cf. Loschky et al., 2014). 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

Boundedness and event cognition

Most studies on event cognition have typically used event segmentation measures to individuate events 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 conceptually coherent interpretations that incorporate the internal temporal contour of the events (i.e., boundedness) and use this information to process continuous streams of visual input. In Experiments 1 and 2, we placed disruptions at different time points during bounded (non-homogenous) and unbounded (homogeneous) naturalistic events and measured the accuracy of detecting these disruptions. The results of both experiments showed that the placement of disruptions affected detection performance for bounded events; for unbounded events, the effect was smaller (Experiment 1) or non-existent (Experiment 2). In Experiment 3, 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. Throughout, these patterns arose even though participants did not have to pay attention to the specific content of the events to perform the task. Thus our hypothesis was confirmed: viewers spontaneously track the temporal texture of events as they make sense of incoming, dynamically unfolding event information.

The present results break new ground in studies of event cognition. First, they show that people compute boundedness during online event comprehension (as opposed to a later process based on 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; Hard et al., 2019; 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. In both respects, our conclusions comport with but go considerably beyond available evidence from explicit event categorization tasks about the role of boundedness in event representation (e.g., Ji & Papafragou, 2020a, 2020b; cf. also Strickland et al., 2015).

Based on the present data, we propose that boundedness should be integrated into existing models of event cognition. One possible path, as suggested already, would be to recruit and enhance the mechanisms outlined in Event Segmentation Theory (Swallow et al., 2009; Zacks et al., 2007). Recall that, according to this theory, viewers predict what is going to happen next in the input stream; furthermore, event boundaries coincide with moments of prediction error brought about by significant changes in event features. 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 since they are equally predictable. By contrast, during bounded events, different temporal slices represent different stages of development.[5]

By turning the zoom lens of the theory towards the internal temporal texture of individual events as opposed to the transition moments between events, our approach expands the scope of current cognitive models of events in several ways. First, it allows us to capture powerful intuitions about the nature of any and every event, including how event representations unfold even before a boundary has been reached, and how otherwise dissimilar event tokens might be grouped into event-general types (e.g., fix a car, fold up a handkerchief and draw a balloon are similar because of their boundedness signatures). Second, it goes beyond a single notion of event

[5] This reasoning is reminiscent of findings that illustrate the role of prediction in the psycholinguistics literature. For instance, word recognition is more impaired when word-initial segments are mispronounced than when word-final segments are mispronounced – presumably because at later time-points people have retrieved the word and do not attend as much to the input (Connine et al., 1993; Marslen-Wilson & Zwitserlood, 1989). Additionally, recent event-related potential evidence indicates that people selectively attend to word onsets only when onset identity is unpredictable; when a word onset can be predicted from the context, it is not treated different from the later parts of the word (Astheimer & Sanders, 2011).

boundary to distinguish between boundaries that coincide with the culmination of an internally articulated event (e.g., fold up a handkerchief) and those that simply coincide with a breakpoint of a structure-less entity (e.g., wave a handkerchief). Third, this approach connects naturally to how events of different types are encoded in language (see next section).

Even though not the focus of our study, it remains highly likely that the present bounded and unbounded events would be treated as distinct units in a segmentation study if they were presented as part of a longer continuous sequence composed of multiple episodes. For instance, a situation in which a child pats a teddy bear, moves a toy car back and forth, and then stacks blocks may be perceived as a sequence of three unbounded events. If in the same situation the car is moved into a toy garage instead, the event sequence would be unbounded-bounded-unbounded. In the EST framework, there would be three event breakpoints as the child starts a new action in both sequences. However, our findings suggest that the three breakpoints would mean different things depending on whether an individual event simply stops (patting a teddy bear, stacking blocks) or culminates at a natural endpoint (moving a toy car into the toy garage).

Finally, it should be noted that in our events, change was tracked as affectedness of an object: many bounded events involve pronounced object changes (e.g., a picture of a balloon comes into being in the event of drawing a balloon). The idea that object state changes play a critical role for event cognition is consistent with findings from an eye tracking study showing that people paid more attention to the action and the affected object at the video offset in events that involved a salient change of state of an object (e.g., peel a potato) compared to events that did not result in a pronounced change (e.g., stir in a pan; Sakarias & Flecken, 2019). This idea also connects with a recent proposal according to which events are represented as a series of intersecting representations of the objects in them (Altmann & Ekves, 2019). Our own approach underscores that what counts as a change in an object over time can be subtle and depend on the viewers' perspective: for instance, even when both bounded and unbounded events involve a perceivable change in the object that is involved in the action (eat a pretzel vs. eat cheerios), the conceptualization of the object as an individual or a non-individuated entity may lead to different assessments of the extent of the change and different boundedness profiles for the event.

Boundedness and the language/cognition interface

Our results offer a way of integrating cognitive theories of events with the notion of boundedness that originated in a long linguistic and philosophical discussion of events, thereby highlighting a homology between event language and event cognition (see also Folli & Harley, 2006; Malaia, 2014; Papafragou, 2015; Strickland et al., 2015; Tversky et al., 2011; Truswell, 2019; for different perspectives, see Takac & Knott, 2015). Our results also cohere with the fact that linguistic boundedness is likely to be a semantic universal (Strickland et al., 2015; von Fintel & Matthewson, 2008), even though its specific instantiations vary across linguistic systems (Bar-el et al., 2005; Botne, 2003). Most broadly, our findings are consistent with evidence that cognition spontaneously extracts other types of event information required for language such as event roles (Hafri, et al., 2013; Hafri, et al., 2018), causality (Kominsky et al., 2017; Leslie & Keeble, 1987; Rolfs, et al., 2013; Wolff, 2007), and animacy (Newman et al., 2010; van Buren et al., 2016), among others. We take the position that, in both our own and these past studies, the structure and content of conceptual event representations form the foundation for linguistic event encoding.

The fact that boundedness is spontaneously computed during event comprehension is particularly useful for language production since the internal temporal profile of an event contributes to the message that the speaker has in mind and wants to talk about. For instance, whether the speaker conceptualizes an action as a continuous, homogeneous (hence unbounded) or a discrete, non-homogeneous (hence bounded) occurrence has consequences for selecting a predicate to describe the action (e.g., *stir* vs. *mix*) and combining the predicate with different types of aspectual markers cross-linguistically (Ferretti et al., 2007; Flecken et al., 2015; von Stutterheim et al., 2012). Similarly, treating an event as having or lacking an inherent boundary can scaffold the way learners acquire the tools for encoding boundedness in their native tongue (van Hout, 2007, 2016, 2018; Wagner, 2012).

Even though the present stimuli were created to be unambiguously bounded or unbounded, it is important to bear in mind that, 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 interact with the agent's preferences, goals and other aspects of the context (Abusch, 1986; Depraetere, 2007; Filip, 2001; Kennedy & Levin, 2008; Mathis & Papafragou, 2022; Zacks & Swallow, 2007). For instance, even though the action of warming a soup does not have a clearly defined endpoint, it is often construed as culminating at the point at which the soup has reached someone's favorite temperature. In expanding on the present work, future 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.

Concluding remarks

We have argued that human cognition spontaneously computes the boundedness profile of an event. Our account integrates insights from the event segmentation literature with more specific representational analyses of event content that arise from the linguistic and philosophical tradition of studying events. We take boundedness to be fundamental to representing temporal entities (just like objecthood is fundamental for representing spatial entities; Bach, 1986) and expect the boundedness profile of an event to have further downstream cognitive consequences for how an event is mentally processed, remembered and described in language.

Acknowledgements

The authors thank Raevyn Johnson for helping construct the videos, and Bryce Lamers, Daniel Bresnahan and Ariel Mathis for helping with collecting and entering data.

Funding

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 #2041171 (A.P.).

References

Abusch, D. (1986). *Verbs of change, causation and time*. [Technical Report CSLI-86-50]. Stanford University: Center for the Study of Language and Information.

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review, 126*(6), 817-840. https://doi.org/10.1037/rev0000154

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia, 49*(12)*,* 3512–3516. https://doi.org/10.1016/j.neuropsychologia.2011.08.014

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. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy, 9*(1)*,* 5–16. www.jstor.org/stable/25001229

Barr, D. J. (2008). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4,* Article 328.

https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bar-El, L., Davis, H., & Matthewson, L. (2005). On non-culminating accomplishments. In L. Bateman, & C. Ussery (Eds.), *Proceedings of the 35th Annual Meeting of the North Eastern Linguistics Society* (Volume 1, pp. 87-102). GLSA.

Botne, R. (2003). To die across languages: Towards a typology of achievement verbs. *Linguistic Typology, 7*(2), 233-278. https://doi.org/10.1515/lity.2003.016

Carruthers, P. (2017). Mindreading in adults: evaluating two-systems views. *Synthese, 194*(3), 673–688. https://doi.org/10.1007/s11229-015-0792-3

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

Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*(2), 193–210. https://doi.org/10.1006/jmla.1993.1011

Csibra, G., Biro, S., Koos, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*(1), 111–133. https://doi.org/10.1207/s15516709cog2701_4

Depraetere, I. (2007). (A)telicity and intentionality. *Linguistics, 45*(2), 243–269. https://doi.org/10.1515/LING.2007.008

Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Kluwer.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 182-196. https://doi.org/10.1037/0278-7393.33.1.182

Filip, H. (2001). Nominal and verbal semantic structure: analogies and interactions. *Language Sciences, 23*(4), 453–501. https://doi.org/10.1016/S0388-0001(00)00033-4

Filip, H. (2004). The telicity parameter revisited. In R. Young (Ed.), *Proceedings of the 14th Semantics and Linguistic Theory Conference* (pp.92–109). LSA.

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

Flecken, M., Gerwien, J., Carroll, M., & von. Stutterheim, C. (2015). Analyzing gaze allocation during language planning: a cross-linguistic study on dynamic events. *Language and Cognition, 7*(1), 138–166. https://doi.org/10.1017/langcog.2014.20

Folli, R., & Harley, H. (2006). What language says about the psychology of events. *Trends in Cognitive Science, 10*(3), 91–92. https://doi.org/10.1016/j.tics.2006.01.002

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880–905. https://doi.org/10.1037/a0030045

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. https://doi.org/10.1016/j.cognition.2018.02.011

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition, 47,* 17–32. https://doi.org/10.3758/s13421-018-0847-z

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586–604. https://doi.org/10.1037/a0024310

Hinrichs, E. (1985). *A compositional semantics for Aktionsarten and NP reference in English* [Unpublished doctoral dissertation]. The Ohio State University.

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. https://doi.org/10.1016/j.jarmac.2017.01.003

Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition, 20*(7)*,* 848–864. https://doi.org/10.1080/13506285.2012.705359

Jackendoff, R. (1991). Parts and boundaries. *Cognition, 41*(1–3), 9–45. https://doi.org/10.1016/0010-0277(91)90031-X

Jackendoff, R. (2007). *Language, consciousness, culture: Essays on mental structure*. Cambridge, MA: MIT Press.

Ji, Y., & Papafragou, A. (2020a). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition, 197,* 104197. https://doi.org/10.1016/j.cognition.2020.104197

Ji, Y., & Papafragou, A (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience*, *35*(10), 1465-1479. https://doi.org/10.1080/23273798.2020.1797839

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* (pp. 156-182). Oxford: Oxford University Press.

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science, 28*(11), 1649-1662. https://doi.org/10.1177/0956797617719930

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, pp. 75-115). Dordrecht: Foris Publications.

Krifka, M. (1998). The origins of telicity. In S. Rothstein (ed.), *Events and grammar* (pp. 197-235). Dordrecht: Kluwer.

Kuhn, J., Geraci, C., Schlenker, P., & Strickland, B. (2021). Boundaries in space and time: Iconic biases across modalities. *Cognition, 210,* 104596. https://doi.org/10.1016/j.cognition.2021.104596

Lakusta, L., & Landau, B. (2005). Starting at the end: the importance of goals in spatial language. *Cognition, 96*(1), 1–33. https://doi.org/10.1016/j.cognition.2004.03.009

Lakusta, L., & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal. *Cognitive Science*, *36*(3), 517–544. https://doi.org/10.1111/j.1551-6709.2011.01220.x

Lee, S. H., & Kaiser, E. (2021). Does hitting the window break it?: Investigating effects of discourse-level and verb-level information in guiding object state representations. *Language, Cognition, and Neuroscience, 36*(8), 921-940. https://doi.org/10.1080/23273798.2021.1896013

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Levin, D. T., & Varakin, D., A. (2004). No pause for a brief disruption: Failures of visual awareness during ongoing events. *Consciousness and Cognition, 13*, 363-372. https://doi.org/10.1016/j.concog.2003.12.001

Loschky, L. C., Ringer R. V., Johnson, A. P., Larson, A. M., Neider, M., & Kramer, A. F. (2014). Blur detection is unaffected by cognitive load. *Visual Cognition, 22*(3), 522-547. https://dx.doi.org/10.1080/13506285.2014.884203

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. https://doi.org/10.3389/fpsyg.2015.01171

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*(5)*,* 533–545. https://doi.org/10.1002/acp.724

Malaia, E. (2014). It still isn't over: Event boundaries in language and perception. *Language and Linguistics Compass, 8*(3), 89–98. https://doi.org/10.1111/lnc3.12071

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. https://doi.org/10.1016/j.neuroimage.2011.10.034

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576

Mathis, A., & Papafragou, A. (2022). Agents' goals affect construal of event endpoints. Submitted.

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* (pp. 27-48). New York: Springer.

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences of the United States of America, 107*(40), 17140–17145. https://doi.org/10.1073/pnas.0914056107

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*(5)*,* 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of*

Personality and Social Psychology, 35(12), 847–862. https://doi.org/10.1037/0022-3514.35.12.847

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. https://doi.org/10.1177/1747021820942479

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science, 34*(6), 1064–1092. https://doi.org/10.1111/j.1551-6709.2010.01107.x

Papafragou, A. (2015). The representation of events in language and cognition. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 327–345). Cambridge, MA: MIT Press.

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. https://doi.org/10.1186/s41235-018-0151-2

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

Pettijohn, K. A., & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation. *Memory & Cognition, 44*(7), 1064-1075. https://doi.org/10.3758/s13421-016-0619-6

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

Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science, 31*(4)*,* 705–719. https://doi.org/10.1080/15326900701399954

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology, 23*(3)*,* 250-254. https://doi.org/10.1016/j.cub.2012.12.017

Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science, 43*(1)*,* 1–30. https://doi.org/10.1111/cogs.12708

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18*(1), 37–55. https://doi.org/10.1002/acp.940

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. https://doi.org/10.1073/pnas.1423080112

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distorts memory in a matter of seconds. *Cognition, 121*(3)*,* 409–415. https://doi.org/10.1016/j.cognition.2011.04.007

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.

https://doi.org/10.1037/a0015631

Takac, M., & Knott, A. (2015). A neural network model of episode representations in working memory. *Cognitive Computation, 7,* 509-525. https://doi.org/10.1007/s12559-015-9330-3

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

Truswell, R., (2019). Event Composition and Event Individuation. In R. Truswell (Ed.), *The Oxford Handbook of Event Structure* (pp. 98–134). Oxford: Oxford University Press.

Tversky, B., Zacks, J., Morrison, J., & Hard, B. (2011). Talking about events. In J. Bohnemeyer & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 216–227). Cambridge: Cambridge University Press.

van Buren, B., Uddenberg, S., & Scholl, B. (2016). The automaticity of perceiving animacy: Goal-directed motion in simple shapes influences visuomotor behavior even when task-irrelevant. *Psychonomic Bulletin & Review, 23*, 797–802. https://doi.org/10.3758/s13423-015-0966-5

van Hout, A. (2007). Acquiring telicity cross-linguistically: On the acquisition of telicity entailments associated with transitivity. In P. Brown, & M. Bowerman (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 255–278). New York: Routledge.

van Hout, A. (2016). Lexical and grammatical aspect. In J. Lidz, W. Synder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics* (pp. 587-610). Oxford: Oxford University Press.

van Hout, A. (2018). On the acquisition of event culmination. In K. Syrett, & S. Arunachalam (Eds.), *Semantics in language acquisition* (pp. 96–121). Philadelphia: John Benjamins Publishers.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review, 66*(2), 143–160. doi: 10.2307/2182371

von Fintel, K., & Matthewson, L. (2008). Universals in semantics. *The Linguistic Review, 25*(1-2)*,* 139–201. https://doi.org/10.1515/TLIR.2008.004

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization: insights from linguistic analysis, eye tracking data and memory performance. *Linguistics*, 50(4), 833−867

Wagner, L. (2012). First Language Acquisition. In Binnick, R. (Ed.), *The Oxford Handbook of Tense and Aspect* (pp. 458–480). Oxford: Oxford University Press.

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition, 90*(2)*,* 163–191. https://doi.org/10.1016/S0010-0277(03)00143-4

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* (pp. 1185–1191). 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, pp. 183–212). Oxford: Oxford University Press.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*(1), 82–111. https://doi.org/10.1037/0096-3445.136.1.82

Zacks, J., Swallow, K., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research, 1076*(1), 150-162. https://doi.org/10.1016/j.brainres.2005.12.122

Zacks, J., Speer, N., Swallow, K., Braver, T., & Reynolds, J. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133* (2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J., & Swallow, M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16,* 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

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

Appendix

Table A.1

Mean accuracy for each bounded event item in Experiment 1.

	Bounded events	Boundedness Source	Accuracy (proportion of correct responses)			
			Mid-break	End-break	No break	Total
1	fold up a handkerchief	Nature of Action	1	0.5	0.875	0.8125
2	put up one's hair		1	0.5	0.9375	0.84375
3	stack a deck of cards		1	0.625	0.9375	0.875
4	group pawns based on color		0.875	0.625	0.9375	0.84375
5	dress a teddy bear		1	0.625	0.9375	0.875
6	roll up a towel		1	0.5	0.9375	0.84375
7	fill a glass with milk		0.875	0.625	0.9375	0.84375
8	scoop up yogurt		1	0.75	0.9375	0.90625
9	draw a balloon	Nature of Affected Object	0.875	0.625	0.875	0.8125
10	tie a knot		1	0.75	0.9375	0.90625
11	eat a pretzel		1	0.625	0.9375	0.875
12	flip a postcard		1	0.625	0.875	0.84375
13	peel a banana		1	0.5	0.875	0.8125
14	blow a balloon		1	0.625	0.9375	0.875
15	tear a paper towel		1	0.5	0.875	0.8125
16	paint a star		0.875	0.625	0.9375	0.84375

Table A.2

Mean accuracy for each unbounded event item in Experiment 1.

	Unbounded events	Boundedness Source	Accuracy (proportion of correct responses)			
			Mid-break	End-break	No break	Total
1	wave a handkerchief	Nature of Action	1	0.875	0.9375	0.9375
2	scratch one's hair		1	0.75	1	0.9375
3	shuffle a deck of cards		1	0.875	0.875	0.90625
4	mix pawns of 2 colors		0.875	0.875	0.9375	0.90625
5	pat a teddy bear		1	0.875	1	0.96875
6	twist a towel		1	0.875	0.9375	0.9375
7	shake a bottle of milk		1	0.875	1	0.96875
8	stir yogurt		1	1	0.9375	0.96875
9	draw circles	Nature of Affected Object	1	1	0.875	0.9375
10	tie knots		0.875	1	0.9375	0.9375
11	eat cheerios		1	0.875	0.9375	0.9375
12	flip pages		1	0.875	1	0.96875
13	crack peanuts		1	0.875	1	0.96875
14	blow bubbles		1	1	1	1
15	tear slices off paper towels		1	0.875	0.9375	0.9375
16	paint stuff		1	0.875	1	0.96875