



Cognitive Science 48 (2024) e13476

© 2024 The Author(s). *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13476

# Aspectual Processing Shifts Visual Event Apprehension

Uğurcan Vurgun,<sup>a</sup> Yue Ji,<sup>b</sup> Anna Papafragou<sup>a</sup>

<sup>a</sup>*Department of Linguistics, University of Pennsylvania*

<sup>b</sup>*School of Foreign Languages, Beijing Institute of Technology*

Received 17 March 2023; received in revised form 29 May 2024; accepted 4 June 2024

---

## Abstract

What is the relationship between language and event cognition? Past work has suggested that linguistic/aspectual distinctions encoding the internal temporal profile of events map onto nonlinguistic event representations. Here, we use a novel visual detection task to directly test the hypothesis that processing telic versus atelic sentences (e.g., “Ebony folded a napkin in 10 seconds” vs. “Ebony did some folding for 10 seconds”) can influence whether the very same visual event is processed as containing distinct temporal stages including a well-defined endpoint or lacking such structure, respectively. In two experiments, we show that processing (a)telicity in language shifts how people later construe the temporal structure of identical visual stimuli. We conclude that event construals are malleable representations that can align with the linguistic framing of events.

*Keywords:* Event; Boundedness; Aspect; Telicity; Endpoints

---

## 1. Introduction

To make sense of the world surrounding them, people divide the continuous flow of input into meaningful segments called *events*. Different sources of information can be used to represent event units (Elman, 2009; Zacks & Swallow, 2007, Zwaan & Radvansky, 1998), including both bottom-up, perceptual features (e.g., changes to an object, or other event participants; Altmann & Ekves, 2019; Lee & Kaiser, 2021; Magliano, Dijkstra, & Zwaan,

---

Data can be accessed at [https://osf.io/5qcpk/?view\\_only=c8a842b1cc2a4f59a7625c7eb1b611a2](https://osf.io/5qcpk/?view_only=c8a842b1cc2a4f59a7625c7eb1b611a2)

Correspondence should be sent to Uğurcan Vurgun, Department of Linguistics, University of Pennsylvania, N 3401 Walnut St., Suite 300C, Philadelphia, PA 19104, USA. E-mail: uvurgun@sas.upenn.edu

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

2001; Newton, Engquist, & Bois, 1977; Sakarias & Flecken, 2019; Zacks, Speer, Swallow, Braver, & Reynolds, 2009) and top-down, conceptual cues (e.g., an agent's goals; Mathis & Papafragou, 2022; Zacks, 2004; cf. Newton, 1973; Vallacher & Wagner, 1987; Wilder, 1978). Although the role of perceptual cues that people observe in events has been defined in detail, less is known about the abstract cues that have top-down effects in event construal. Here, we investigate whether linguistic event descriptions function as a top-down cue and affect the way viewers apprehend visual events in predictable ways.

There are currently only a handful of studies addressing the relationship between language and visual event processing. In an early eye-tracking study (Papafragou, Hulbert, & Trueswell, 2008), English- and Greek-speaking adults were found to inspect motion events differently when asked to verbally describe them, thereby reflecting the effects of language-specific speech planning on eye gaze patterns; however, the eye gaze differences disappeared when participants were simply asked to look at each event to prepare for later memory questions. In a more recent study, foregrounding an agent or a patient in a picture description task was found to influence initial attention allocation for the following event (Saupe & Flecken, 2021). Furthermore, in offline tasks, overt language has been known to produce framing effects (Tversky & Kahneman, 1981) with downstream implications for event memory (Faber & Gennari, 2015; Fausey & Boroditsky, 2010; Skordos et al., 2020, Wang & Gennari, 2019). In all these senses, language can act as a zoom lens on event construal (Gleitman, 1990).

In the current study, we focus on whether and how prior linguistic information about an event could change the temporal structure-building (also known as *boundedness*; Ji & Papafragou, 2020a) during the dynamic representation of the event. Boundedness is a foundational part of event apprehension, and evidence that it aligns flexibly with linguistic information can illuminate the degree of malleability in event construals, as well as the kinds of cues that can be used to mentally assemble events. To preface our experimental findings, we give some background on boundedness and its role in event cognition before we turn to relevant distinctions in language.

### 1.1. *Boundedness in event cognition*

Event boundedness (Ji & Papafragou, 2020a, 2020b) is defined as the presence of an inherent endpoint in the mental representation of an event: bounded events (e.g., someone writing a letter) include distinct temporal stages and a well-defined endpoint, while unbounded events lack a distinct endpoint (e.g., someone writing). Much recent work shows that the bounded versus unbounded distinction characterizes how viewers process events (Ji & Papafragou, 2020a, 2020b; Malaia, 2014; Papafragou & Ji, 2023; Strickland et al., 2015; Wehry, Hafri, & Trueswell, 2019; Wellwood, Hespos, & Rips, 2018). For instance, when exposed to videos of everyday visual events, viewers learn to place stimuli such as folding a napkin, blowing a balloon, eating a pretzel, and dressing a teddy bear into one ("bounded") category, and stimuli such as waving a napkin, blowing bubbles, eating cheerios, and patting a teddy bear into another ("unbounded") category, and are able to extend these abstract categories to new stimuli (Ji & Papafragou, 2020a). This ability is present already in young children

(Ji & Papafragou, 2020b) and persists even when participants are prevented from describing events verbally through a linguistic shadowing task in which participants are asked to count forward by 2 from the number they saw on the screen until they encounter a new number (Ji & Papafragou, 2020a). Thus, the bounded-unbounded distinction captures an aspect of event representations that is active independently of language.

Crucially, boundedness is computed in real-time during online event apprehension, even when it is not relevant to the task. In a recent study (Ji & Papafragou, 2022), participants watched short video clips of events that were known to be biased toward either a bounded construal (e.g., blowing a balloon) or an unbounded construal (e.g., blowing bubbles). Video clips in some cases included brief visual interruptions placed at either the midpoint (50% of the timeline) or a late point (80% of the timeline) of the event. Participants had to detect whether there was an interruption in the video clip. For bounded events, viewers were more likely to miss brief visual interruptions placed at late-points compared to midpoints of events, but no such difference emerged for unbounded events. The authors argue that, for bounded events, whose internal texture has distinct substages and leads to the highly informative moment of culmination, what happens later is deemed to be important and attracts processing resources compared to what happens in event midpoints: for these events, irrelevant interruptions are neglected when they appear close to the event endpoint compared to the midpoint (on the prominence of endpoints, see Lakusta & Landau, 2005; Papafragou, 2010; Pettijohn & Radvansky, 2016; Regier & Zheng, 2007; Strickland & Keil, 2011; Swallow, Zacks, & Abrams, 2009; Zheng & Goldin-Meadow, 2002). By contrast, for unbounded events, whose temporal texture lacks well-defined endpoints, there is little or no difference in the detection of interruptions across the event timeline. According to the paradigm, the ability to detect interruptions can serve as an indicator of how visual events are being processed. This method builds on the well-established idea that, during moments when a visual stimulus requires more processing resources, less attention is available for external distractors irrelevant to the event, and the accuracy of detecting (content-irrelevant) interruptions decreases (Huff, Papenmeier, & Zacks, 2012; Papenmeier, Maurer, & Huff, 2019; cf. also Mack, 2003; Mack & Rock, 1998, Simons, 2000).

## 1.2. Linguistic aspect and event boundaries

In language, distinctions homologous to boundedness that pertain to the temporal profile of events are encoded via lexical aspect (also studied as *predicational aspect* or *telicity*; Egg, 2020; Filip, 1993; Folli & Harley, 2006; Krifka, 1992, 1998; van Hout, 2016). A *telic* sentence depicts developments toward a “climax” (Vendler, 1957), “culmination” (Parsons, 1990), or a “terminal point” (Comrie, 1976). For instance, (1a) describes a situation with a beginning, a midpoint, and an endpoint. By contrast, *atelic* sentences depict a structure without an inherent endpoint (Hinrichs, 1985; Krifka, 1989, 1998; Taylor, 1977): in (1b) and (1c), the occurrence lacks a specific endpoint, and, in principle, could go on forever.

- (1) a. The girl wrote a letter. (telic)
- b. The girl wrote letters. (atelic)
- c. The girl did some writing. (atelic)

Telic sentences are compatible with time-span adverbials, while atelic sentences can canonically be modified by durative adverbials, as shown in (2).

- (2) a. The girl wrote a letter in 30 minutes. (telic)  
 b. The girl wrote letters/did some writing for 30 minutes. (atelic)

Telicity is a compositional interpretation of the temporal profile of events influenced by multiple elements in a sentence, including verbs and noun phrases—as in (1) (Bach, 1986; Champollion, 2017; De Swart, 1998; Jackendoff, 1991, 1997; Kamp & Reyle, 1993; Krifka, 1989; Moens & Steedman, 1988; Pustejovsky & Bouillon, 1995; Talmy, 1978; van Lambalgen & Hamm, 2005; Verkuyll, 1972, 1993) but also adverbial phrases, verb particles, prefixes, prepositional phrases, and even contextual elements (Brinton, 1985; Filip, 1993; van Hout, 1996; Jackendoff, 1997; Moens & Steedman, 1987; Pustejovsky, 1991, 1995). Sometimes, the addition of linguistic material can change the canonical interpretation of a sentence (Brennan & Pyllkänen, 2008; Jackendoff, 1997; Moens & Steedman, 1987; Piñango, Zurif, & Jackendoff, 1999; Pustejovsky, 1991, 1995; Todorova, Straub, Badecker, & Frank, 2000): while (1a) is a telic sentence, the interpretation shifts to an atelic one when a durative adverbial is introduced as in (3a). Changes in the opposite direction (atelic to telic) are also possible, as shown by comparing (1b) to (3b).

- (3) a. The girl wrote a letter for an hour. (atelic)  
 b. The girl wrote letters in two hours. (telic)

It has long been observed in the linguistic literature that the same event can be described with both telic and atelic sentences (as in the examples (1a) and (1c) above). Thus, physical entities in the eventuality do not unambiguously determine telicity (Bennett & Partee, 1972); rather, viewers may describe comparable scenes with telic or atelic sentences depending on how they interpret the event (e.g., *She is running a mile* vs. *She is exercising*). Correspondingly, the telicity of the description may reveal information about how the speaker viewed and interpreted the event.

Is the opposite effect possible? Can an aspectual description change how a viewer interprets an event? Some studies have explored this question focusing on grammatical aspect (or “outer aspect”) that is related to how one perceives the internal time structure of a situation. This distinction is closely related to lexical aspect (or “inner aspect”) since both features overlay the event time with multiple layers of temporal structure, thereby delineating its temporal attributes (van Hout, 2016). Consider the distinction between perfective and imperfective aspect in English (“She wrote a letter” vs. “She was writing a letter”; van Hout, 2016). In the former, the speaker views the situation as an indivisible entity, while in the latter, the situation is represented as it is unfolding, highlighting the internal structure. Unlike English, German lacks a grammatical encoding of the imperfective aspect. Flecken, Carroll, Weimar, and Von Stutterheim (2015) found that German speakers attended more to the endpoints of motion events compared to English speakers, even in an entirely nonverbal context (see also von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012). Similarly, in a self-paced reading study that compared perfective and imperfective aspect in description sentences, Madden-Lombardi and colleagues showed that people had longer response times

when pictures related to the events were not congruent with grammatical aspect (e.g., an in-use corkscrew picture in a sentence with perfective aspect; Madden-Lombardi, Dominey, & Ventre-Dominey, 2017).

Turning to prior research on how telicity interfaces with event cognition, the few existing studies have exclusively used either linguistic responses or reading times and have, therefore, not directly examined visual event processing. For instance, Wagner and Carey (2003) found that when participants were asked questions about an event with a telic verb phrase (“How many times does the boy blow up the balloon?”), they used different counting strategies compared to cases in which the questions included atelic sentences (“How many times does the boy blow?”; cf. Barner & Snedeker, 2005). It remains an open question whether lexical aspect effects on event processing could be revealed using nonlinguistic (visual) measures.

### 1.3. *Current study*

In this study, we ask whether telicity in language could guide boundedness or the temporal structure of events computed during visual event apprehension. Specifically, in two experiments, we probe whether a telic versus atelic description shown prior to an event might influence the temporal profile of the event as measured by a newly developed visual detection task that has revealed a signature of (un)boundedness (Ji & Papafragou, 2022). We also compare results to a case where no description was offered prior to the event. Of interest is whether the specific aspectual perspective in the linguistic input could switch the temporal structure of the event that would arise from simple observation of the visual stimulus.

The outcome of this empirical investigation is theoretically important for three reasons. First, aspectual processing effects on event apprehension would be significant for theories of event cognition. Recall that according to prominent event models, observers use several visual cues such as object state changes to understand how an event unfolds and when it ends (Altmann & Ekves, 2019; Lee & Kaiser, 2021; Sakarias & Flecken, 2019). However, most of these features are dependent on how much information observers have within their reach because what counts as a change or what could be considered an endpoint is rarely obvious (cf. Ji & Papafragou, 2022). By probing whether linguistic information about the temporal profile of events could be used to overcome specific biases in the perceptual input, this study adds to a growing literature on top-down effects on event understanding. This is significant because linguistic input may supply aspectual information that can influence the apprehension of what counts as change within an event.

Second, and relatedly, the outcome of the present experiments bears on the nature of boundedness within event cognition. In prior studies that examined how people understand the temporal profile of events, participants were presented with stimuli that were constructed so as to be readily perceived as either having or lacking a boundary (Ji & Papafragou, 2020a, 2020b, 2022; Strickland et al., 2015; Wellwood et al., 2018). As a result, the malleability of (un)boundedness construals has not been explored, even though several commentators have acknowledged that the same visual input might often be construed as either a bounded or an unbounded event (e.g., writing a letter vs. writing; see Ji & Papafragou, 2020a, 2022). The present study asks how the (linguistic) perspective of the viewer might organize a temporally

unfolding stream of sensory information; as such, it highlights the viewer's active role in computing whether an event unfolds toward a specified endpoint or not.

Third, our study has significant implications for models of the language-cognition interface. Several studies have proposed that aspect connects to temporal properties of event cognition (Filip, 1993; Folli & Harley, 2006; Ji & Papafragou, 2020a, 2022; Malaia, 2014; Wellwood et al., 2018) but the present study is unique in explicitly bringing the two phenomena together by testing how aspectual sentence processing interfaces with event understanding in the visual world. To the extent that aspect frames event apprehension, the current study would support a rich theoretical model of event structure that includes a correspondence between psychological and linguistic perspectives on events.

## 2. Experiment 1

In Experiment 1, we asked whether (a)telicity in linguistic descriptions could affect how people process the temporal profile of events. On each trial, participants read a background story ending in either a canonically telic or an atelic sentence, and then watched a short video clip of an event that, on the basis of prior work (e.g., Ji & Papafragou, 2020a, 2020b), was known to be biased toward a bounded construal (e.g., was taken to have a well-defined endpoint; e.g., fold a napkin). Video clips in some cases included brief visual interruptions placed at either the midpoint (50% of the timeline) or a late point (80% of the timeline) of the event. Participants had two tasks after watching the video: (a) determine whether the earlier sentence was compatible with the video (primary task), (b) indicate whether there was a glitch (interruption) in the video clip (secondary task). Since in critical trials, the sentences were designed to match the video, of interest was mostly the answer to the “glitch” question and its relation to the exposure sentence. In a control condition, participants were not exposed to a sentence and only answered this second question.

Recall that in recent work adopting this paradigm (Ji & Papafragou, 2020b, 2022), when viewers construed events as bounded in a nonlinguistic task, they were more likely to miss brief visual interruptions placed at late-points compared to midpoints of events, but no such difference emerged for other events construed as unbounded. In the present experiment, we used differential sensitivity to the placement of visual interruptions as an index of whether *the same event* was construed as bounded or unbounded. We hypothesized that the telic versus atelic description shown prior to the video clips might influence participants' construal and detection of interruptions at different time points of the videos. Specifically, we predicted that telic descriptions would leave the bounded profile of the events unaffected—hence the signature preference to attend to what happens at endpoints compared to midpoints for these events (and the tendency to miss late-occurring interruptions) should persist. Not providing any description prior to the event at all (cf. the control condition) should lead to a similar pattern. We further expected that an atelic description shown prior to the video clips would shift viewers' preferred construal such that the event would now be processed as an unbounded one, and mid- and late interruptions would be detected at similar rates. In sum, we expected an interaction between the aspectual profile of an event description and the placement of an

interruption. Therefore, the main idea in the current study is that linguistic input (i.e., telic or atelic) provides a template for the event representation and this template could overcome the inherent biases within the visual input and impose an event construal without an inherent endpoint.

## 2.1. Method

### 2.1.1. Participants

We recruited 180 speakers through the online recruitment platform Prolific. Participants reported being monolingual English speakers and were compensated for their time at an hourly rate of \$9.

Using data from a pilot study, we did a power analysis to find the smallest sample size for 80% statistical power with the significance level set at 5%. The factor of interest was a significant interaction between condition (telic vs. atelic) and the placement of interruptions (mid vs. late). The detection rates for midpoint and late-point interruptions were computed in each condition. Based on the detection rate difference between the midpoint and late-point interruptions, our power analysis showed that the smallest sample size was 110 people for an effect size of 0.58 (Cohen's  $d$ ). The final, larger sample was based on the addition of a control condition (see next section).

### 2.1.2. Stimuli and procedure

Fifteen video clips (taken from Ji & Papafragou, 2020a) were used in this experiment ( $M = 10.3$  s,  $SD = 2.3$  s, see Part A in Supporting Information). All of the videos included the same woman performing an action that gradually resulted in an object's change of state. Each of the videos was considered in principle compatible with two target verb phrases, one telic and the other atelic (e.g., *draw a balloon* vs. *do some drawing*, or *fold a napkin* vs. *do some folding*). However, prior norming studies (Ji & Papafragou, 2020a) had confirmed that, in the absence of other information, these video clips were classified as having "a beginning, a midpoint, and an endpoint," and were described with telic sentences 97.7% of the time. In other words, these stimuli were typically treated as bounded events.

We edited each video clip in DaVinci Resolve at the time point corresponding to the midpoint (50% of the event timeline) or a late point (80% of the event timeline). In the editing process, we removed one frame (approximately 0.03 s) at the corresponding time points (the videos had a display rate of 30 frames per second) to create a momentary visual interruption (or "glitch"). In general, removing one frame is enough for people to detect the removal (Ji & Papafragou, 2022; Shady, MacLeod, & Fisher, 2004). At the end of editing, each video clip had three versions (an interruption at the midpoint, an interruption at the late point, or no interruption). For an approximation of the editing result, see Fig. 1.

The experiment was run on PCIBex (Zehr & Schwarz, 2018), and implemented online on the recruitment platform Prolific. Participants were randomly assigned to one of three between-subject conditions: Telic, Atelic, and No Sentence. At the beginning of the session for the Telic and Atelic conditions, participants saw a picture of the woman in the videos and read the following scenario: "This is Ebony. She has just recovered from orthopedic surgery.



Fig. 1. Examples of two interruption types for the event of folding a napkin: a midpoint interruption (top) and a late-point interruption (bottom). The arrows indicate the placement of the “glitch” (30 ms).

Now she needs some extra help with her fine motor movements and coordination. Ebony’s physical therapist gave her a set of timed exercises with household objects to determine how she is doing now. Your task is to watch the videos and see whether she did the exercise. There may be some glitches in the videos because of Ebony’s camera. After each video, please also let us know if you notice a glitch.”

At the beginning of each trial in the Telic and Atelic conditions, participants saw the exercise formulated as either a telic or an atelic sentence describing what Ebony should do. A sample pair of sentences for the event of Fig. 1 is shown in (4).

- (4) a. Ebony should fold a napkin in 9 seconds. (telic)
- b. Ebony should do some folding for 9 seconds. (atelic)

All telic sentences included change of state verbs (e.g., *fold*, *stack*) with a quantified noun phrase (e.g., *a napkin*, *a deck of cards*) that served as the incremental theme: there was always a homomorphism between the affected object and the time course of the denoted event (Dowty, 1991; Krifka, 1989), such that the changes in the object tracked or “measured out” the way the event developed toward an inherent endpoint (Tenny, 1987). The telic sentences included a time-span modifier that further delimited the endpoint of the event (e.g., *in 9 seconds*). The atelic sentences involved the light verb construction *do some V-ing* using the same V as the telic sentences but in mass syntax (e.g., *some folding*) to describe a continuous activity (Barner, Wagner, & Snedeker, 2008; Wellwood et al., 2018). There was no noun phrase denoting the affected object. The atelic sentences included a durative modifier that carved out a portion out of the continuous activity (e.g., *for 9 seconds*). The time length mentioned in both sentences (e.g., 9 seconds) always reflected the actual duration of each stimulus. (All events and sentences are given in Part A of Supporting Information). Sentences were shown in the center of the screen for 6.5 s. Then, the sentence disappeared, and the video started.

Each participant saw a total of 15 videos. For nine (critical) items, sentences in both conditions matched the content shown in the video. In the remaining six videos (fillers), the sentences did not match the action in the video (e.g., the description was about *bouncing* a balloon, and the actor in the video clip *blew* the balloon). There were equal numbers of trials ( $n = 5$ ) with midpoint interruptions, late-point interruptions, or no interruptions (in each case, three of the five trials were critical items, and two were fillers). Each participant saw only one of the edited versions of each video clip. Each participant saw an equal number of trials ( $N =$



5) with each type of interruption (midpoint, late point, or no interruptions). After each video, participants saw the first question: “Ebony’s time is up. Did she do the exercise?” and had to choose from the “Yes” and “No” answer options. Even though ostensibly the primary task was to check whether the actor did the exercise, the true question of interest was the next one: “Any glitch in the video?”. Again, participants had to select a “Yes” versus “No” answer.

In the No Sentence condition, the video clips were shown without any sentences. Participants were asked to see whether there is a glitch in the video, so they only performed the interruption detection task. The whole session lasted 9 minutes on average in the Telic and Atelic conditions, and 6 minutes on average in the No Sentence condition.

## 2.2. Results

The following methods have been used for both Experiment 1 and Experiment 2. Binary accuracy data were analyzed using mixed-effects modeling (Baayen, Davidson, & Bates, 2008). All models were fitted using the *glmer* function of the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2021). The contrasts in factors were computed with the *emmeans* function in the *emmeans* package in R (Lenth, 2021), and the results were corrected for multiple comparisons with the Tukey method. Categorical predictors were coded using sum contrasts. Random intercepts were added to the model for each Subject and Item (Baayen et al., 2008; Barr, 2008). The fixed-effects and random-effects structure of our linear mixed-effects models was determined using a stepwise model comparison approach. Initially, we started with a simple model including only random intercepts for each participant (ID). We incrementally added random intercepts for other factors (e.g., Items) and systematically tested the inclusion of fixed effects and their interactions (e.g., Condition, Interruption). Each subsequent model was compared against the previous ones using likelihood ratio tests, examining changes in AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) to assess improvements in model fit without overfitting. This iterative process allowed us to identify the most appropriate random-effects structure that accounts for both subject and item variability while ensuring model stability and convergence.

Overall, participants were overwhelmingly accurate when answering the question of whether Ebony had done the exercise or not in both the Telic ( $M = 99.8\%$ ,  $SE = 0.003$ ) and Atelic ( $M = 98.5\%$ ,  $SE = 0.013$ ) conditions (with a significant difference between the two conditions,  $\chi^2(1) = 13.31$ ,  $p = .0003$ , presumably reflecting the bias in the visual stimuli toward bounded construals). This is important because it showed that descriptions of either telicity profile matched the videos as assessed by high accuracy rates for the verification question. Similarly, accuracy on filler items (where the sentences were not compatible with videos) for this question was not different across conditions (Telic:  $M = 88\%$ ,  $SE = 0.09$ , Atelic:  $M = 90\%$ ,  $SE = 0.07$ ) ( $\chi^2(1) = 0.59$ ,  $p = .44$ ). Therefore, we focus on data from the interruption detection task (“Any glitch in the video?”) across conditions.

In our main analysis, we examined the Mid versus Late points (Interruption) interruption detection accuracy on critical items across the Telic, Atelic, and No Sentence (Condition) conditions, as shown in Fig. 2. For the analysis, factors were coded with sum contrasts. We submitted the binary accuracy data to a mixed model with fixed effects of Condition (Telic,

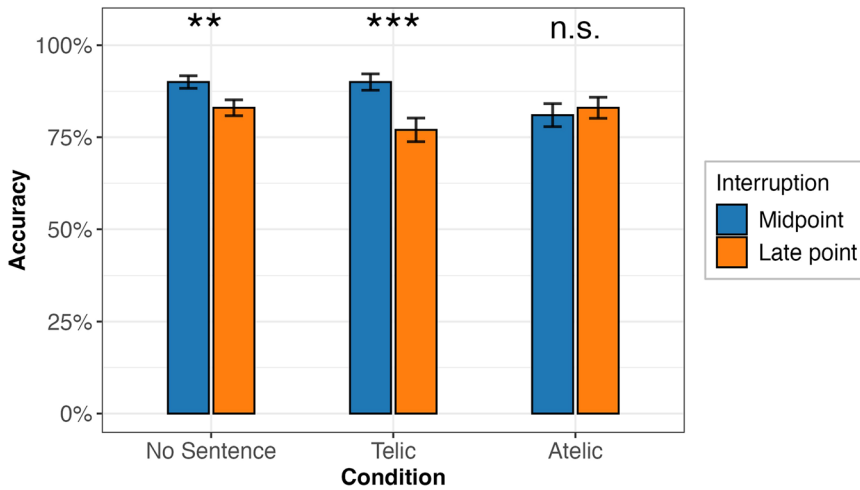


Fig. 2. Proportion of correct responses in critical trials of Experiment 1. Error bars represent  $\pm$  SEM.

Table 1

Fixed effect estimates for the mixed effects model of accuracy on critical trials in Experiment 1

Effect	$\chi^2$	<i>df</i>	<i>z</i>	<i>p</i> value
Condition (No Sentence, Telic, Atelic)	1.40	2	1.18	.50
Interruption Type (Midpoint vs. Late point)	14.13	1	3.76	< .001***
Condition * Interruption Type	6.83	2	2.61	.033**

*Note.* Formula in R:  $\text{Acc} \sim 1 + (1|\text{Subject}) + (1|\text{Item}) + \text{Condition} + \text{Interruption Type} + \text{Condition} : \text{Interruption Type}$ . Asterisks indicate levels of statistical significance: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

Atelic, No Sentence), Interruption Type (Midpoint, Late point), and their interaction (see Table 1). Condition did not affect overall interruption detection accuracy (Telic:  $M = 84\%$ ,  $SE = 0.02$ , Atelic:  $M = 82\%$ ,  $SE = 0.02$ , No Sentence:  $M = 87\%$ ,  $SE = 0.01$ ) ( $\chi^2(2) = 1.404$ ,  $p = .4956$ ). There was a significant effect of Interruption Type, such that participants detected midpoint interruptions ( $M = 88\%$ ,  $SE = 0.01$ ) significantly better than late-point interruptions ( $M = 81\%$ ,  $SE = 0.02$ ) ( $\chi^2(1) = 14.13$ ,  $p < .001$ ). Crucially, there was a significant interaction between Condition and Interruption Type ( $\chi^2(2) = 6.83$ ,  $p = .033$ ). Follow-up analyses showed that, as predicted, participants in the No Sentence condition performed significantly worse in trials with Late-point interruptions ( $M = 83\%$ ,  $SE = 0.04$ ) compared to Midpoint interruptions ( $M = 90\%$ ,  $SE = 0.03$ ) (*odds ratio* = 2.09,  $SE = 0.6$ ,  $p = .01$ ). Similarly, participants in the Telic condition performed significantly worse in trials with Late-point interruptions ( $M = 77\%$ ,  $SE = 0.04$ ) compared to Midpoint interruptions ( $M = 90\%$ ,  $SE = 0.03$ ) (*odds ratio* = 4.41,  $SE = 1.72$ ,  $p = .0001$ ). However, participants in the Atelic condition responded similarly in trials with Late point ( $M = 83\%$ ,  $SE = 0.04$ ) and Midpoint ( $M = 81\%$ ,  $SE = 0.04$ ) interruptions (*odds ratio* = 1.11,  $SE = 0.4$ ,  $p = .77$ ).

### 2.3. *Di Note. scussion*

In this experiment, we examined whether verifying (a)telic sentences shown prior to events affected how participants construed the boundaries of events (as measured by their detection of unrelated interruptions along the event timeline). Since all videos depicted (what was preferentially viewed as) bounded events containing inherent endpoints, in the absence of other information, viewers' attention was drawn to endpoints to the detriment of detecting irrelevant or external features of the stimuli (such as unrelated interruptions) placed near endpoints (see No Sentence condition). Similarly, in the Telic condition, participants were more likely to miss irrelevant late-point interruptions compared to midpoint interruptions. However, in the Atelic condition, having to verify atelic sentences about the upcoming video clip led them to process the occurrence as an unbounded event, resulting in similar detection rates for both interruption types. Thus, having to verify description sentences shown prior to video clips changed how people process visual event information as the event unfolded.

Notice that the present patterns could not be explained by the possibility that participants found the atelic stimulus harder to process (even if they accepted it as a potential descriptor of the video), and were thus less focused on its content and more attentive to glitch detection. If participants were distracted by the atelic sentences (compared to the telic ones), we would see a global decrease in accuracy from the Telic to the Atelic condition. The results of Experiment 1, however, show no such difference. Furthermore, this alternative explanation does not capture the interaction between Condition and Interruption Type, which is at the heart of our results (and is predicted by our own account). Therefore, it is unlikely that the patterns are driven by an overall preference for the telic over the atelic descriptions.

## 3. Experiment 2

Why did the sentences in Experiment 1 change how people processed visual input to form event construals? One possibility is that, as we have assumed, the results are driven by the difference between the aspectual profile of the sentences in the Telic and Atelic conditions. However, it is also possible that the presence of a direct object Noun Phrase (NP) (e.g., *a teddy bear*, *a napkin*) in the Telic but not the Atelic condition simply increased attention to these objects and their changing state on a surface level. The aspectual influence of direct objects has been shown in several studies (Dowty, 1991; Krifka, 1992). Thus, the disregard for endpoints when people were exposed to Atelic sentences could be (partially or completely) due to the absence of these concrete objects (e.g., *do some drawing*). To exclude this possibility, participants in Experiment 2 were exposed to minimally different sentences across the two telicity manipulations. Experiment 2 also included a wider variety of events compared to Experiment 1.

### 3.1. *Methods*

#### 3.1.1. *Participants*

We recruited 180 speakers through the online recruitment platform Prolific. Participants reported being monolingual English speakers and were compensated for their time at an hourly rate of \$9.

Table 2

Fixed effect estimates for the mixed effects model of accuracy on critical trials in Experiment 2

Effect	$\chi^2$	df	z	p value
Condition (No Sentence, Telic, Atelic)	11.21	2	3.35	.004**
Interruption Type (Midpoint vs. Late point)	0.018	1	0.13	.90
Condition * Interruption Type	11.05	2	3.32	.004**

Note. Formula in R:  $\text{Acc} \sim 1 + (1|\text{Subject}) + (1|\text{Item}) + \text{Condition} + \text{Interruption Type} + \text{Condition: Interruption Type}$ . Asterisks indicate levels of statistical significance: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

### 3.1.2. Stimuli and procedure

Stimuli and procedure were similar to Experiment 1 with two changes. First, we revised the sentences such that now both telic and atelic sentences included a direct object noun phrase and only differed in the adverbial phrases indicating (a)telicity (Brinton, 1985; Filip, 1993; Van Hout, 1996; Jackendoff, 1997; Moens & Steedman, 1987; Pustejovsky, 1991, 1995). As before, the telic sentences included a time-span adverbial (e.g., *in 10 seconds*), as shown in (5a). Atelic sentences included a durative adverbial (e.g., *for 10 seconds*) that served to shift the eventual interpretation into an atelic one, as shown in (5b).

- (5) a. Ebony should draw a balloon in 10 seconds. (telic)
- b. Ebony should draw a balloon for 10 seconds. (atelic)

Second, using materials from Ji and Papafragou (2020a), we expanded the stimulus set to include 21 items per participant (as opposed to 15 items per participant in Experiment 1) to increase empirical coverage. For 15 (critical) items, the sentences in both conditions matched the content shown in the video clips. In the remaining six items (fillers), the sentences did not match the action in the video clips. (See Part B in Supporting Information for a list of all items.) Each participant saw seven midpoint, seven late point, and seven no interruption trials, with five trials of each type for critical items, and two for filler items.

### 3.2. Results

Overall, participants were overwhelmingly correct when answering the initial question of whether Ebony had done the exercise or not in both the Telic ( $M = 99.9\%$ ,  $SE = 0.001$ ) and Atelic ( $M = 99.5\%$ ,  $SE = 0.006$ ) conditions ( $\chi^2(1) = 3.28$ ,  $p = .07$ , a marginal difference). The high accuracy rates for the verification question confirmed that descriptions of either telicity profile matched the video. Therefore, we focus on data from the interruption detection task across conditions (“Any glitch in the video?”).

In our main analysis, we examined the interruption detection accuracy rates for mid versus late points on critical items across the Telic, Atelic, and No Sentence conditions. Performance is shown in Fig. 3. We submitted the binary accuracy data to a mixed model with fixed effects of Condition (Telic, Atelic, No Sentence), Interruption Type (Midpoint, Late point), and their interaction (see Table 2). Condition had a significant effect on overall interruption detection accuracy (Telic:  $M = 67\%$ ,  $SE = 0.02$ , Atelic:  $M = 62\%$ ,  $SE = 0.02$ ,

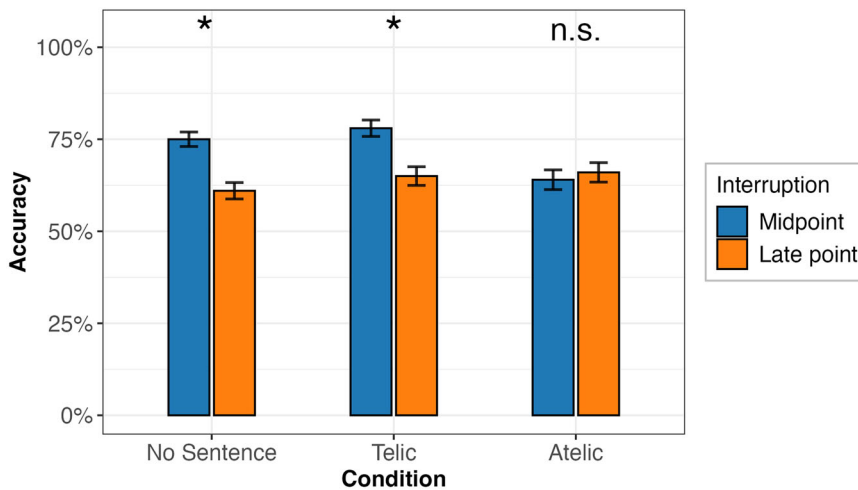


Fig. 3. Proportion of correct responses in the critical trials of Experiment 2. Error bars represent  $\pm$  SEM.

No Sentence:  $M = 68\%$ ,  $SE = 0.02$  ( $\chi^2(2) = 11.21$ ,  $p = .003678$ ). There was no difference between the Telic and Atelic ( $odds\ ratio = 1.65$ ,  $SE = 0.498$ ,  $p = .2247$ ) or the Telic and No Sentence condition ( $odds\ ratio = 1.63$ ,  $SE = 0.5$ ,  $p = .2486$ ) but performance in the No Sentence condition was significantly better than in the Atelic condition ( $odds\ ratio = 2.69$ ,  $SE = 0.769$ ,  $p = .0016$ ). Interruption Type did not have a significant effect on interruption detection accuracy: participants detected midpoint interruptions and late-point interruptions similarly ( $\chi^2(1) = .02$ ,  $p = .895$ ).

Crucially for present purposes, there was a significant interaction between Condition and Interruption Type ( $\chi^2(2) = 11.05$ ,  $p = .00399$ ). Follow-up analyses showed that, as predicted, participants in the No Sentence condition performed significantly worse in trials with Late-point interruptions ( $M = 61\%$ ,  $SE = 0.02$ ) compared to Midpoint interruptions ( $M = 75\%$ ,  $SE = 0.02$ ) ( $odds\ ratio = 0.618$ ,  $SE = 0.137$ ,  $p = .0299$ ). Similarly, participants in the Telic condition performed significantly worse in trials with Late-point interruptions ( $M = 65\%$ ,  $SE = 0.03$ ) compared to ones with Midpoint interruptions ( $M = 78\%$ ,  $SE = 0.02$ ) ( $odds\ ratio = 1.76$ ,  $SE = 0.39$ ,  $p = .0104$ ). However, participants in the Atelic condition responded similarly in trials with Late point ( $M = 66\%$ ,  $SE = 0.03$ ) and Midpoint ( $M = 64\%$ ,  $SE = 0.03$ ) interruptions ( $odds\ ratio = 0.95$ ,  $SE = 0.2$ ,  $p = .812$ ).

### 3.3. Discussion

In Experiment 2, we used minimally different sentences for the Telic and Atelic conditions to remove an alternative interpretation of the data from Experiment 1. We also expanded the event repertoire compared to the earlier study. The results replicated the major finding from Experiment 1. When participants watched bounded events (e.g., events typically taken to have a distinct endpoint) after reading telic sentences, their interruption detection patterns were not uniform across different time points (i.e., they performed significantly worse in tri-

als with Late point compared to Midpoint interruptions). However, when participants watched the same events after reading atelic sentences, their midpoint versus late-point detection patterns were similar. Therefore, linguistic input shown prior to the videos affected how people processed dynamic perceptual input.<sup>1</sup>

## 4. General discussion

Events are foundational for human cognition and have a critical role for representing, remembering, and understanding the world. To form event representations, people combine both perceptual (Magliano et al., 2001; Newton et al., 1977; Zacks, 2004; Zacks et al., 2009) and nonperceptual cues (Mathis & Papafragou, 2022; Newton, 1973; Vallacher & Wagner, 1987; Wilder, 1978; Zacks, 2004). In the current study, we examined whether language could function as a valid cue that would introduce a perspective for processing events. In two experiments, viewers watched short movie clips depicting bounded events and were asked to indicate whether they saw an interruption that occurred at either the midpoint or the endpoint of the events. We found that telic versus atelic sentences shown prior to the movie clips influenced event construal and hence the detection of interruptions at midpoints versus endpoints. Specifically, when the aspectual profile of the sentence mismatched the construal arising from the visual input alone (as in the Atelic condition), aspectual information would overturn the specific bias in the visual input. In other words, the linguistic input worked as a template for event viewers to process the temporal structure of unfolding events: the aspectual description gave viewers an idea of what would be relevant for the purposes of apprehending the event and thus focused their attention to the internal time points within the event in different ways. This novel finding provides direct experimental evidence for the role of lexical aspect in event apprehension and supports a correspondence between lexical aspect in language and temporal structure-building in event apprehension.

### 4.1. *Language and the malleability of event representations*

The present findings contribute to current event cognition theories by specifying the role of higher-order linguistic information in perceiving and understanding events. Recall that observers use several visual cues such as object state changes to understand how an event unfolds and when it ends (Altmann & Ekves, 2019; Lee & Kaiser, 2021; Sakarias & Flecken, 2019). For instance, an eye-tracking study has shown 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). In a recent proposal, an event is defined as a series of intersecting representations of the object(s) in the event (Altmann & Ekves, 2019). These studies suggest that a salient change in the affected object can be perceived as a clear event boundary and thus lead to a bounded construal. Our findings, however, indicate that a physical change in an object does not uniquely determine whether an event is interpreted as bounded; what counts as a change instead depends on the viewers' perspective. For

instance, even when there is a perceivable change in an object that is involved in the action (*fold a napkin*), a language-driven conceptualization of the event (*do some folding*) may lead to an unbounded construal. These results show that event cognition is a flexible process that depends on and integrates various sources of input (see Mathis & Papafragou, 2022, for further discussion).

The malleability of event construals is noteworthy in the context of prior research on event boundedness. As mentioned in the Introduction, prior studies that investigated how viewers draw a distinction between bounded and unbounded events used event stimuli that were designed to be readily perceived as belonging to one of the two categories (Ji & Papafragou, 2020a, 2020b, 2022). The current work provides the first empirical evidence for a conclusion that prior work only hinted at, namely, that boundedness is in the eye of the beholder and can shift even for the very same visual stimulus. In this position, viewers act as invisible directors who impose a perspective on the visual world, and seemingly neutral expressions such as “observe” fail to describe what people do when they are exposed to dynamic visual information.

More broadly, our results cohere with a line of work suggesting that the process of producing or comprehending linguistic information can frame our understanding of an event by acting as a “zoom lens” on some aspects of the event over others (Gleitman, 1990; Johnson, Raye, Wang, & Taylor, 1979; Mathis & Papafragou, 2022; Papafragou et al., 2008; Sauppe & Flecken, 2021) and suggest a broader alignment between linguistic and cognitive representations (Jackendoff, 1983; Landau & Jackendoff, 1993, see also Ünal, Ji, & Papafragou, 2021 for a review).

#### 4.2. *Aspect and event structure*


Linguistic research on aspect has long hypothesized that the distinction between telicity and atelicity may extend beyond the language domain (Filip, 1993; Folli & Harley, 2006; Malaia, 2014; Shipley & Zacks, 2008; Wellwood et al., 2018). Recent research has identified such a possible correlate in the cognitive domain—the notion of boundedness defined in terms of the presence of well-defined event endpoints (Ji & Papafragou, 2020a, 2022; Papafragou & Ji, 2023; cf. Kuhn et al., 2021; Strickland et al., 2015). Going beyond previous work, our current data establish a direct link between (a)telicity in language and (un)boundedness in cognition by showing that (a)telicity introduces a perspective on the temporal profile of what could be the very same stream of experience. Even a small change in the use of temporal adverbials (as in Experiment 2) in linguistic event descriptions can trigger different readings about event temporal structure, which can further change the way people process dynamic input. An important advance compared to past work is the use of a visual detection task to probe how the verification of aspectual utterances interfaces with systems of event cognition. This new paradigm expands the tools on the large linguistic and psycholinguistic literature on how aspect is understood (Zwaan & Radvansky, 1998; Zwaan, Langston, & Graesser, 1995; van Hout, 2016) and offers a precise and independently verified way of probing the comprehender’s commitments during aspectual processing.

Several questions remain ripe for further investigation of how the aspectual framing of a sentence influences how observers perceive the temporal profile of an unfolding event. First, it is important to test whether (as expected) aspectual effects would work in the opposite direction, that is, whether viewers' prior exposure to telic sentences would shift unbounded event construals to bounded ones. Second, given the well-known variability in how aspect is encoded across languages (Filip, 2008; Kagan, 2010; Mittwoch, 2019), it is important to move beyond English and ask whether aspectual sentences cross-linguistically would create similar effects.

## Acknowledgments

This research was supported by NSF BCS grant # 2041171 to A.P.

## Open Research Badges

 This article has earned an Open Data Badge. Data is available at [https://osf.io/5qcpk/?view\\_only=c8a842b1cc2a4f59a7625c7eb1b611a2](https://osf.io/5qcpk/?view_only=c8a842b1cc2a4f59a7625c7eb1b611a2).

## Note

1 A careful reader might notice that participants appear to perform worse overall in the detection task in Experiment 2 compared to Experiment 1 (see Figs. 2 and 3). This difference holds even when we only restrict attention to the very same items used in both experiments. We do not have an explanation for this phenomenon, but we suspect that it is linked to variability sometimes observed in virtually recruited samples over periods of time (Chandler, Mueller, & Paolacci, 2014; Higgins, McGrath, & Moretto, 2010; Komarov, Reinecke, & Gajos, 2013).

## References

- Altmann, G., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, 126(6), 817–840.
- 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: Linguistic structure and causation*. Oxford, England: Oxford University Press.
- Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*, 97(1), 41–66.
- Barner, D., Wagner, L., & Snedeker, J. (2008). Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns. *Cognition*, 106(2), 805–832.
- Barr, D. (2008). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.



- Brennan, J., & Pykkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, 106(2), 132–143.
- Brinton, L. J. (1985). *The development of English aspectual systems: Aspectualizers and post-verbal particles*. Cambridge, England: Cambridge University Press.
- Bennett, M., & Partee, B. (1972). *Toward the logic of tense and aspect in English*. Santa Monica, CA: System Development Corporation.
- Champollion, L. (2017). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Oxford, England: Oxford University Press.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46(11), 112–130.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, England: Cambridge University Press.
- De Swart, H. (1998). *Introduction to natural language semantics*. Stanford, CA: Center for the Study of Language and Information.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547–619.
- Egg, M. (2020). Aspectual composition: “Drinking (a glass of) milk”. In *The Wiley Blackwell Companion to Semantics*.
- Elman, J. L. (2009). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In J. P. de Ruiter (Ed.), *Language and cognition: A collection of research papers on language acquisition and use* (pp. 3–33). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Faber, M., & Gennari, S. P. (2015). In search of lost time: Reconstructing the unfolding of events from memory. *Cognition*, 143, 193–202.
- Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17(5), 644–650.
- Filip, H. (1993). *Aspect, situation types and nominal reference*. University of California, Berkeley dissertation.
- Filip, H. (2008). Events and maximalization. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect* (pp. 217–256). Amsterdam: John Benjamins.
- Flecken, M., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French–German L2 users. *Modern Language Journal*, 99(S1), 100–122.
- Folli, R., & Harley, H. (2006). What language says about the psychology of events. *Trends in Cognitive Sciences*, 10(3), 91–92.
- Gleitman, L. R. (1990). The structural sources of verb meanings. In J. Pustejovsky & S. L. Epstein (Eds.), *Semantics and the lexicon* (pp. 15–56). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Higgins, C., McGrath, E., & Moretto, L. (2010). MTurk crowdsourcing: A viable method for rapid discovery of Arabic nicknames? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk* (pp. 89–92).
- Hinrichs, E. W. (1985). *A compositional semantics for Aktionsarten and NP reference in English (aspect, Montague grammar, events, mass terms)*. Ohio State University dissertation.
- Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, 20(7), 848–864.
- Jackendoff, R. S. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. S. (1991). *Semantic structures*. Cambridge, MA: MIT Press.
- Jackendoff, R. S. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Ji, Y., & Papafragou, A. (2020a). Is there an end in sight? Viewers’ sensitivity to abstract event structure. *Cognition*, 197, 104197.
- Ji, Y., & Papafragou, A. (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience*, 35, 1465–1479.
- Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language*, 127, 104353.

- Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory*, 5(3), 229.
- Kagan, O. (2010). Russian aspect as number in the verbal domain. In B. Laca & P. Hofherr (Eds.), *Layers of aspect* (pp. 125–146). CSLI Publications.
- Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Komarov, S., Reinecke, K., & Gajos, K. Z. (2013). Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 207–216).
- 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* (pp. 75–115). Foris Publications.
- Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In I. Sag & A. Szabolcsi (Eds.), *Lexical matters* (pp. 29–54). Stanford, CA: CSLI 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.
- Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, 96(1), 1–33.
- Landau, B., & Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2), 217–238.
- Lee, S. H. Y., & 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), 921–994.
- Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.10.
- Mack, A. (2003). Inattention blindness: Looking without seeing. *Current Directions in Psychological Science*, 12(5), 180–184.
- Mack, A., & Rock, I. (1998). Inattention blindness: Perception without attention. *Visual Attention*, 8, 55–76.
- Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PLoS One*, 12(12), e0189919.
- Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (2001). Generating predictive inferences while viewing a movie. *Discourse Processes*, 32(2–3), 129–162.
- Malaia, E. (2014). Neural correlates of emotion–cognition interactions: A review of evidence from brain imaging investigations. *Psychology & Neuroscience*, 7(2), 145–157.
- Mathis, A., & Papafragou, A. (2022). Agents’ goals affect construal of event endpoints. *Journal of Memory and Language*, 127, 104373.
- Mittwoch, A. (2019). Aspectual classes. In R. Truswell (Ed.), *The Oxford handbook of event structure* (pp. 29–49). Oxford University Press.
- Moens, M. F., & Steedman, M. J. (1987). *Computational semantics: An introduction to artificial intelligence and natural language understanding*. Cambridge, England: Cambridge University Press.
- Moens, M. F., & Steedman, M. J. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, 14(2), 15–28.
- Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, 34(6), 1064–1092.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38.
- Newton, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35(11), 847–862.

- Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184.
- Papafragou, A., & Ji, Y. (2023). Events and objects are similar cognitive entities. *Cognitive Psychology*, *143*, 101573.
- Papenmeier, F., Maurer, A. E., & Huff, M. (2019). Linguistic information in auditory dynamic events contributes to the detection of fine, not coarse event boundaries. *Advances in Cognitive Psychology*, *15*(1), 30.
- 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.
- Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of Psycholinguistic Research*, *28*, 395–414.
- Pustejovsky, J. (1991). *The generative lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, J. (1995). *The syntax of event structure*. Cambridge, MA: MIT Press.
- Pustejovsky, J., & Bouillon, P. (1995). *Lexical semantics: The problem of polysemy*. Oxford, England: Clarendon Press.
- R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria.
- Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science*, *31*(4), 705–719.
- Sakariäs, 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), e12708.
- Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition*, *206*, 104516.
- Shady, S., MacLeod, D. I., & Fisher, H. S. (2004). Adaptation from invisible flicker. *Proceedings of the National Academy of Sciences*, *101*(14), 5170–5173.
- Shipley, T. F., & Zacks, J. M. (Eds.). (2008). *Understanding events: From perception to action*. Oxford University Press.
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155.
- Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology*, *37*(5–6), 254–270.
- Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–415.
- 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*, *112*(19), 5968–5973.
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236.
- Talmy, L. (1978). Figure and ground in complex sentences. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 4, pp. 625–649). Stanford, CA: Stanford University Press.
- Taylor, J. R. (1977). Linguistic categorization: Prototypes in linguistic theory. *Linguistic Inquiry*, *8*(2), 369–422.
- Tenny, C. L. (1987). *Grammaticalizing aspect and affectedness*. Massachusetts Institute of Technology.
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In L. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.
- Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science*, *13*(1), 224–242.

- Vallacher, R. R., & Wagner, G. P. (1987). Cognitive consistency and attitude change: The role of intra-individual processes. *Journal of Personality and Social Psychology*, 53(6), 1092–1100.
- van Hout, A. (2016). Lexical and grammatical aspect. In J. J. Lidz, W. Synder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 587–610). Oxford University Press.
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, 66(2), 143–160.
- Verkuyl, H. J. (1972). *On the compositional nature of the aspects*. Dordrecht, Netherlands: Reidel.
- Verkuyl, H. J. (1993). *A theory of aspectuality: The interaction between temporal and atemporal structure*. Cambridge, England: Cambridge University Press.
- van Hout, A. (1996). *Universal grammar and language learnability*. Oxford, England: Blackwell Publishers.
- van Lambalgen, M., & Hamm, F. (2005). *The proper treatment of events*. Oxford, England: Blackwell Publishing.
- von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, 50(4), 833–867.
- Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition*, 90(2), 163–191.
- Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. *Cognitive Psychology*, 108, 1–21.
- 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* (Vol. 2) (pp. 183–212). Oxford University Press.
- Wilder, D. A. (1978). Effect of predictability on units of perception and attribution. *Personality and Social Psychology Bulletin*, 4(2), 281–284.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979–1008.
- Zacks, J., & Swallow, M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80–84.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2009). Event perception: A mind-brain perspective. *Psychological Bulletin*, 135(1), 74–118.
- Zehr, J., & Schwarz, F. (2018). PennController for internet-based experiments (IBEX). <https://doi.org/10.17605/OSF.IO/MD832>
- Zheng, M., & Goldin-Meadow, S. (2002). Thought before language: How deaf and hearing children express motion events across cultures. *Cognition*, 85(2), 145–175.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6(5), 292–297.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendices