Agents' goals affect construal of event endpoints

Ariel Mathis & Anna Papafragou

University of Pennsylvania

Author Note

Ariel Mathis & Anna Papafragou, Department of Linguistics, University of Pennsylvania

Correspondence concerning this article should be addressed to Ariel Mathis, Department of Linguistics, University of Pennsylvania at 3401 Walnut St., Suite 300C, Philadelphia, PA, 19104.

Contact: apmathis22@gmail.com; (901) 412-2389

Abstract

Theories of event cognition have hypothesized that the boundaries of events are characterized by change, including a change in the agent's goal, but the role of higher-order goal information in how people conceptualize events is currently not well-understood. In a series of experiments, we used a novel method to test whether goals can affect how viewers determine when an event ends. Participants read a context sentence stating an agent's goal (e.g., "Jesse wants to eat the orange for her breakfast", "Jesse wants to use the orange as a garnish"). Participants then saw an image of a partly complete visual outcome (e.g., a partly peeled orange) and were asked to identify whether an event had occurred ("Did she peel the orange?"). Participants were more likely to accept a partly complete outcome if the outcome satisfied the agent's goal (Experiments 1 and 2). This goal effect was present even when participants saw an image that corresponded to a mostly complete visual outcome (e.g., a mostly peeled orange; Experiment 3). Our results offer the first direct evidence in support of the conclusion that higher-order goal information affects the way even simple physical events are conceptualized. The further suggest that theories of event cognition need to account for the rich and varied informational sources used by the human mind to represent events.

Keywords: events, goals, intentionality, context, endpoint

Agents' goals affect construal of event endpoints

Daily life, from morning routines and daily commutes to getting ready for bed at night, can be viewed as a series of events. Events have been characterized as temporal entities unfolding at a specific time and location and having a beginning and ending point (Zacks & Tversky, 2001). Events can consist of smaller event units: "making coffee" can be thought of as an event by itself, but this overarching event can also be broken down into smaller subevents such as inserting the filter, putting in the ground coffee, pouring the water, pressing the "on" button, and the pot filling with brewed coffee (Zacks, Tversky, & Iyer, 2001).

Several models have been proposed to explain how people process events. According to one prominent model (Event Segmentation Theory, or EST; Zacks, Speer, Swallow, Braver, & Reynolds, 2007), event comprehension is an ongoing process that is facilitated by the use of multiple simultaneous event models that are maintained in working memory. Event models are used to predict near future occurrences and are updated when there is an increase in prediction errors. These increases in prediction error, as indicated by transient changes in neural activity, correspond to the placement of boundaries during both active event segmentation and passive viewing (Zacks et al., 2001). An increase in prediction errors, and the corresponding detection of event boundaries, has been found to occur at points of change in the stream of input (Speer, Zacks, & Reynolds, 2004; 2007). These changes can be perceptual such as changes in movement (e.g., a car turns; Magliano, Kopp, McNerney, Radvansky, & Zacks, 2012). Additionally, conceptual knowledge about goals and intentions also plays a role in identifying event boundaries (Zacks, 2004). For instance, a stream of actions is divided into smaller units when viewers are uncertain about the goal of making these actions (Newtson, 1973; Vallacher & Wegner, 1987; Wilder, 1978). Relatedly, very young infants can parse everyday actions by

placing boundaries at the points where a goal is achieved (Baldwin, Baird, Saylor, & Clark, 2001; cf. also Brandone & Wellman, 2009; Luo & Johnson, 2009; Woodward, 1998).

Despite the emphasis on how people identify event boundaries, research on event segmentation has not addressed the question of how people process the representational unit *within* event boundaries (i.e., "what happens" within an event, which in turn connects to our conceptual understanding of when an event begins and ends; Ji & Papafragou, 2020a, b). For instance, placing a boundary within a stream of events indicates an event breakpoint but does not indicate whether at that breakpoint the event came to an end (i.e., finished or culminated), or simply stopped. Distinguishing between these options requires a mechanism for tracking moment-by-moment changes within a single event (as opposed to global transitions from one event to another – the traditional focus of EST; Zacks et al., 2007); furthermore, it requires a framework beyond EST for understanding how changes along one or more dimensions of an event support construals of how the event unfolds and ends (cf. Kurby & Zacks, 2012; see also Huff, Meitz, & Papenmeier, 2014; Zacks, Baily, & Kurby, 2018).

In an attempt to address the internal temporal structure of individual events, a more recent theory of event representation known as Intersecting Object Histories (IOH; Altmann & Ekves, 2019) argues that, in addition to a beginning and ending point, an event must also contain a change of state in some object. Within this framework, events are defined as intersecting histories of objects undergoing a change of state. Importantly, according to IOH, the change in an object state occurs independently of any observer. The idea that object state changes play a critical role for event cognition receives support from neuroscientific data on how people represent events. For instance, when processing verbal descriptions of events where an object undergoes a change of state ("He chopped an onion, then he smelled the onion"), adults track the

causal history of the object, as evidenced by the fact that the pre- and post-change states of the object (here, the chopped and the intact onion) appear to be in competition (no such competition is experienced when actions are performed on different object tokens, as in "He chopped an onion, then he smelled another onion"; Hindy, Altmann, Kalenik, & Thomspon-Schill, 2012; Solomon, Hindy, Altmann, & Thompson-Schill, 2015; see also Kang, Eerland, Joergensen, Zwaan & Altmann, 2019; Misersky, Slivac, Hagoort & Flecken, 2021). In further suggestive eye tracking findings, people pay more attention to the action and the affected object at the video offset in events that produce a salient change of state for an object (e.g., peel a potato) compared to events that do not result in a pronounced change (e.g., stir in a pan; Sakarias & Flecken, 2019; see also Lee & Kaiser, 2021). Similarly, people remember ceased events that induce an object state change better than events that do not produce such a change (Santin, van Hout & Flecken, 2021). IOH captures the intuition that events can be taken to end (or culminate) when the object that is affected by the event reaches a complete change of state; it therefore provides a way of incrementally tracking how an event unfolds through gradual object change. It remains open, however, whether information other than object-state transformations can be tied to event culmination.

Intentionality in event representation

A key but understudied component of event representation in the line of research just reviewed concerns the role of higher-order factors for the way events are structured. The event segmentation literature recognizes that both perceptual and conceptual/intentional cues contribute to event boundaries (and thus event conceptualization). However, the roles of these two types of cues have typically been intertwined in existing paradigms, with conceptual/

intentional changes often corresponding to perceptual changes (Tversky, Zacks, & Hard, 2008; Zacks, 2004; cf. also Brandone & Wellman, 2009; Luo & Johnson, 2009; Woodward, 1998). For instance, in many of the classic event segmentation studies, it is not possible to completely isolate changes in the goals of an event agent from co-occurring spatiotemporal cues such as character movement as participants view and segment film clips. Consider the event of making a pot of coffee again: a change in the agent's goal (e.g., completing the goal of filling the coffee maker with water and then deciding to turn it on) is also accompanied by a distinct change in movement (the change in motion away from the reservoir towards the 'on' button).[1] Relatedly, within IOH, the emphasis until now has been on links between event structure and accumulated, observer-independent change in an object, not on higher-order considerations bearing on event structure: even though Altmann and Ekves (2019) briefly suggest a role for intentionality and other abstract factors in event representation, this role is limited to situations in which an observer anticipates likely outcomes and states of objects based on perceived goals and intentions. Ideally, to gain a better understanding of how the mind represents events, one needs to be able to introduce and manipulate intentionality information clearly and in isolation from the physical components of an unfolding event (including, among other things, the state of the affected object).

A different tradition that has examined event structure in narrative texts can offer a useful perspective. This work is relevant because both behavioral patterns of event boundary placement (Magliano et al., 2012) and neural activation around event boundaries (Speer et al., 2007; Zacks

[1] One study by Levine, Hirsch-Pasek, Pace, and Golinkoff (2017) attempted to eliminate such spatiotemporal cues in a film segmentation task by playing motion clips in reverse. Participants shown the reversed film continued to segment similarly to those shown the original film. However, as the authors note, while reversing the film reduced the available spatiotemporal cues, the cohesion of the agent's movements was not eliminated.

et al., 2001) have been found to be similar regardless of whether an event is presented visually or in a narrative text. According to one proposal, understanding a narrative involves building situation models (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998), that is, mental representations of the situations described within a narrative. Later developments of this idea suggest that situations in narratives are centered on events: as comprehenders monitor situational continuity, they index, track and update events within situation models along several dimensions, including time, space, protagonists, causation and – importantly for present purposes – intentionality (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998). According to this broad literature, readers maintain mental ongoing lists of characters' goals as they process narratives; these lists are updated as the story progresses, and characters achieve (or abandon) existing goals and add new ones to the list (Bower & Rinck, 1999; Suh & Trabasso, 1993; Magliano & Radvansky, 2001; Trabasso & Wiley, 2005; see also Kopatich, Feller, Kurby, & Magliano, 2019). In what follows, we introduce a novel paradigm combining linguistic/descriptive and visual components to explore the role of intentionality in event cognition.

Current study

In the current study we tested how contextually supplied knowledge about the goals of an agent combines with physical visual cues within an event (here, the state of an object) to affect how viewers construe how an event unfolds and especially how it ends. In Experiment 1, we presented participants with a context sentence that introduced the goal of an agent (e.g., "Jesse wants to eat the orange for her breakfast", or "Jesse wants to use the orange as garnish"). People were then shown images of objects depicting a visual event outcome at a stage of partial

completion (e.g., a partly peeled orange) and had to answer a question about the event (e.g., "Did she peel the orange?"). Critically, the event in the test question was an intermediate step (or subevent) in fulfilling the agent's overarching goal. We were interested in whether participants would be more likely to give non-culmination responses (e.g., to deny that the agent peeled the orange) when the agent's stated goals involved a higher degree of subevent development (as in "eat the orange", that requires that all of the orange be peeled) as opposed to a lower degree of development (as in "use the orange as garnish", where even a small piece of the peel is enough). In subsequent experiments, we further explored the effect of an agent's goals on computations of event endpoints (Experiment 2) and asked whether the effect of goal context extends to cases that present – on visual grounds - mostly complete event outcomes (e.g., a mostly peeled orange; Experiment 3).

The present paradigm contains two methodological innovations. As previously discussed, many of the methods utilized in prior work on event cognition were insufficient or simply not designed to isolate the role of intentionality from that of visual features of the input stream. The solution employed in the current experiments was to use a combination of descriptive text and static images of event outcomes depicting various stages of completion. The use of a partially linguistic format to present goal information allows for the agent's goal to be made explicit while also allowing the manipulation of the goal in isolation from other cues (cf. also Madden & Zwaan, 2003; and previous section on narratives). Similarly, the use of a static image allows for the manipulation of visual cues to event progression and culmination. The choice of static pictures is further justified by evidence that event information can be reliably extracted from a single event snapshot (e.g., Hafri, Papafragou, & Trueswell, 2013; Ünal & Papafragou, 2019; cf. Hindy et al., 2012; Solomon et al., 2015).

Our study bears on current theories of how the mind represents real-world events. Beginning with segmentation accounts, our approach goes 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) to ask how viewers combine different sources of information to understand the conceptual content and organization of *a single* event, including the moment at which the event ends. Recall that EST does not have a mechanism for tracking moment-by-moment changes within individual events but instead focuses on global transitions from one event to another (Zacks et al., 2007). Relatedly, EST aims to explain how one event ends and the next one begins but remains silent about whether the boundary in the right periphery of an event actually represents a true conceptual endpoint (i.e., the moment the action finishes or culminates) or something else (e.g., a point at which the action simply stops, or is interrupted). Our study goes beyond the purview of EST to investigate the hypothesis that viewers integrate unobservable (intentional) alongside observable (visual) cues to update the content of their event representations, including event endings.

Our study also bears on the main claims of IOH. Our paradigm uses as a starting point the robust finding that the canonical development of an event often tracks the transformations of an object affected by the event (peeling an orange tracks the state of the orange, and the natural course of the event is complete when the orange is completely peeled; see Hindy et al., 2012; Solomon et al., 2015, among others). Of interest is whether tracking object transformations (within what we will call Visual Outcomes) offers a unique pathway into event structure, or alternatively should be combined with intentionality information. Of further interest is whether intentionality plays a role when visual information for event completion is ambiguous (e.g., the transformation of the object is only partial, as in Experiments 1 and 2) or also when visual

information is more definitive (e.g., the transformation of the object is mostly complete, as in Experiment 3). If intentionality affects conceptualization of when an event ends, and does so across both ambiguous and more advanced visual changes to an object, the strict focus on an object's history to define and delimit events within IOH would need to be revised. We test these possibilities in the experiments that follow.

Experiment 1

Data Availability

The datasets and Stata analysis codes for this and subsequent experiments in the current paper are available at https://osf.io/udh34/.

Participants

Forty-three native English speakers were recruited from the undergraduate subject pool of the Psychological and Brain Sciences department at the University of Delaware. Participation in the study fulfilled a course requirement. The sample size was based on prior studies of event cognition (Madden & Zwaan, 2003; Ji & Papafragou, 2020a) and was used for setting participant numbers in later experiments.

Stimuli

Visual Stimuli. A total of 54 images were included in Experiment 1. Of these, the 18 items were target images depicting Partly Complete visual outcomes (see Section A of Supplemental Materials for a full list). The remaining 36 were filler items. Filler items consisted of 18 Incomplete and 18 Complete outcome images. All of the images depicted an object at various stages of visual change.

Images for both test and filler items were selected after norming for the presumed degree of event progress based on the visual outcome depicted in the image. The norming study for test

items consisted of 225 images representing 32 events. Images were taken from incremental points along the timeline of each event. Each event was depicted in either five (19 events) or ten (13 events) images depending on the overall length of the action. Eighteen participants were recruited from the Psychological and Brain Sciences department subject pool at the University of Delaware and completed the studies online through Qualtrics Survey Software (Qualtrics, Provo, UT) as part of a course requirement. Participants read a sentence containing an agent's goal (e.g., "Jasmine wants to peel the orange"), and saw an image (i.e., a partly peeled orange). Then they responded to a question about the visually available stage of the temporal unfolding of the event ("What percent did she peel the orange?") using a sliding scale from 0% -100% in increments of 10%. Image presentation was randomized within Qualtrics for each participant. Filler images were normed separately from the target items in a similar way using participants from the same pool (N=25). Table 1 contains the mean norming score and standard deviation for the visual outcomes included in Experiment 1.

Verbal Stimuli. Each image was accompanied by a test question containing a telic Verb Phrase in perfective aspect that required a Yes or No answer (e.g., "Did she peel the orange?", for the target item in Figure 1). The verb phrases used in the study were chosen for their scalar properties or the fact that they denoted incremental changes. The test questions described a sub-event necessary for the agent to complete an overarching goal (i.e., the orange must be peeled to be eaten).

Table 1. Mean norming scores (and standard deviations) by categories of Visual Outcomes included in Experiment 1.

Visual Outcome	'What percent…?'	SD

Incomplete	7.91%	4.76%
Partly Complete	27.02%	7.91%
Complete	92.78%	5.43%

For each filler and target image, we also provided a Context to be displayed before the image. Contexts consisted of a single sentence that stated the event agent's goal (and included either *want* or *need*). For each filler, there was a single Context (Figure 1). For each target (such as the partly peeled orange in Figure 1), there were 3 possible Contexts: (a) High Goal Contexts introduced a goal for which a greater development of the subevent was needed (e.g., "Jesse wants to eat the orange for her breakfast", where the orange needs to be completely or almost completely peeled to be eaten); (b) Low Goal Contexts introduced an overarching goal that could be satisfied even by a relatively modest degree of progress along the subevent timeline (e.g., "Jesse wants to use the orange as a garnish", where a small amount of peeling an orange can

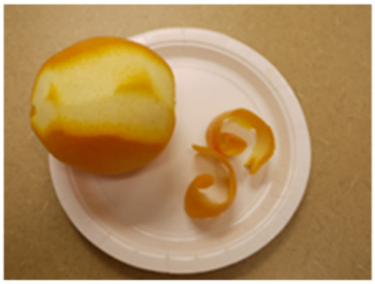
Incomplete
Context: Megan plans to take her lunch to school.

Did she pack the apple?

Complete
Context: Grayson wants to scare his little sister with the balloon.

Did he pop the balloon?

Partly Complete
Low Goal Context: Jesse wants to use the orange as a garnish.
Neutral Goal Context: Jesse wants to peel the orange.
High Goal Context: Jesse wants to eat the orange for her breakfast.



Did she peel the orange?

Figure 1. Sample visual outcomes and verbal prompts (Experiment 1).

yield enough for a garnish); (c) Neutral Goal Contexts simply included a goal later found in the test question ("Jesse wants to peel the orange"; cf. our earlier norming study). See Figure 1 for examples and Section A in Supplemental Materials for a full list.

Procedure

Visual Outcome (Incomplete, Complete, and Partly Complete) and - for Partly Complete Outcomes – Context (Low Goal, Neutral Goal, and High Goal) were within-subjects variables. Three lists were created by counterbalancing the Contexts for Partly Complete Outcomes and participants were assigned to one of the lists. All participants saw a total of 54 trials: 18 involved Incomplete Outcomes, 18 Complete Outcomes, and 18 Partly Complete Outcomes (with 6 Partly Complete Outcomes in each of the 3 types of Context: Low Goal, High Goal, and Neutral Goal). Experiment 1 was programmed and administered in OpenSesame. Trial order was randomized separately for each participant within the OpenSesame software.

Participants were asked to "read the following scenarios, look at the accompanying image, and answer each question" prior to beginning the experiment. Each trial began with a fixation point and participants moved on by pressing the spacebar on the keyboard. The Context sentence was then shown. Participants were instructed to press the spacebar after reading the sentence. Next the image of an event at a certain visual outcome appeared below the Context sentence. The test question and response options (Yes/No) automatically appeared below the image after an additional 500ms. The Context sentence, image, and test question remained on screen until a response was given by pressing "d" for Yes and "k" for No.

Analysis

The data were analyzed separately for responses based on the Visual Outcomes (Incomplete, Partly Complete, Complete) and responses based on the Linguistic Context (Low Goal, Neutral Goal, High Goal) manipulation. Responses for both analyses were coded on a binary scale (Yes = 1; No = 0). Analysis of the responses to the Visual Outcomes was conducted using the entire dataset. The Linguistic Context analysis was conducted on responses to Partly Complete items only (after we collapsed across Contexts). Data were analyzed using the melogit function in Stata version 16.1 (StataCorp, 2019) to perform a multilevel mixed-effects logistic regression. The same data analysis strategy was followed in all further experiments.

Results

Visual Outcomes. The Visual Outcome data for Experiment 1 consisted of 43 participants x 54 items = 2,322 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 2 summarizes the data. The best fit for these data was a model that included Visual Outcome (Incomplete, Partly Complete, Complete) as a first level predictor. Table 2 presents the odds ratios for the multi-level model of Visual Outcome. For this analysis, Partly Complete outcomes were set as the comparison level. Unsurprisingly, Partly Complete outcomes ($M = 0.39$) elicited Yes responses significantly more often than Incomplete outcomes ($M = 0.05$, $p < .001$) and significantly less often than Complete outcomes ($M = 0.85$, $p < .001$). Visual information, therefore, clearly affected the construal of event endpoints.

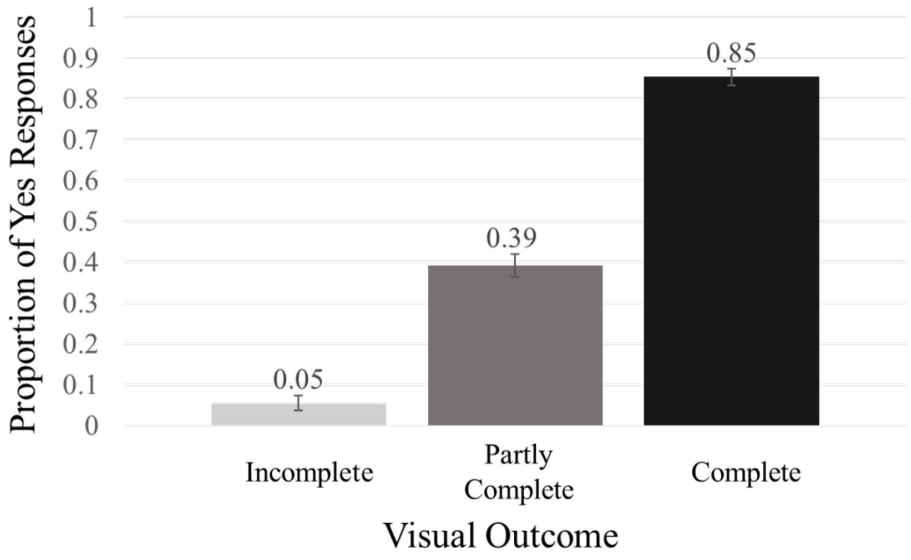


Figure 2. Proportion of Yes responses by Visual Outcomes (all items) in Experiment 1.

Table 2. Odds ratios for the multi-level model of Yes responses by Visual Outcomes (all items) in Experiment 1. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	0.58	0.15	-2.17*
Incomplete	0.07	0.03	-6.76***
Partly Complete	1.00		
Complete	18.22	6.84	7.74***

Linguistic Context. The Linguistic Context data for Experiment 1 consisted of 43 participants x 18 items = 774 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 3 summarizes the data. The best fit for these data was a model that included Context (Low Goal, Neutral Goal, High Goal) as a first level predictor. Table 3 presents the odds ratios for the multi-level model of Context. For this analysis, Neutral Goal was used as the comparison level. Low

Goal contexts elicited Yes responses significantly more often ($M = 0.49$) than Neutral Goal contexts ($M = 0.35$, $p < .001$). There was no difference in responses to Neutral Goal and High Goal contexts ($p = 1.00$).

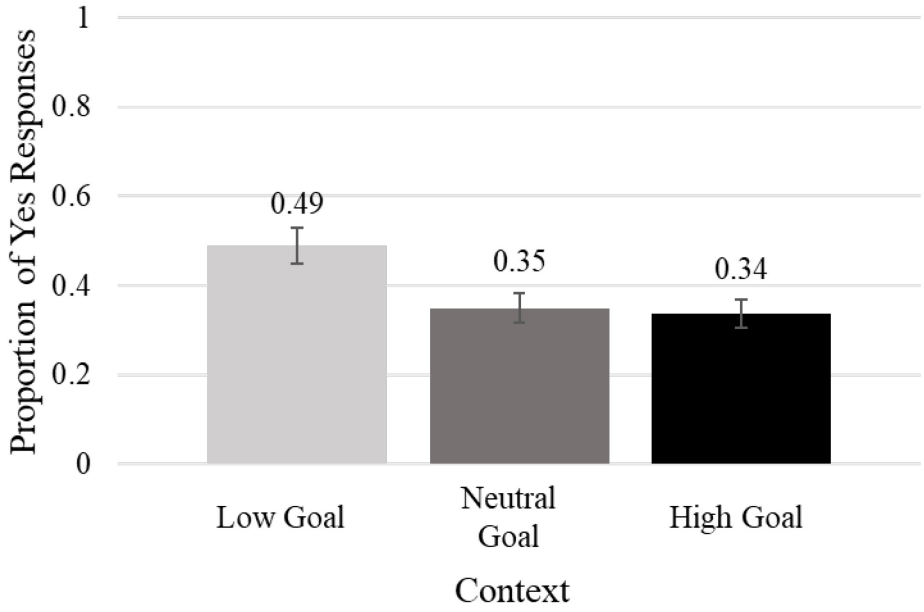


Figure 3. Proportion of Yes responses by Context (Partly Complete Outcomes only) in Experiment 1.

Table 3. Odds ratios for the multi-level model of Yes responses by Context (Partly Complete responses only) in Experiment 1. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	0.41	0.15	-2.46*
Low Goal	2.35	0.52	3.87***
Neutral Goal	1.00		
High Goal	0.95	0.21	-0.21

Discussion

In Experiment 1, we asked whether participants would be more likely to accept that an event occurred if the event outcome satisfied an agent's intention. Participants were informed about an agent's goal (i.e., "Jesse wants to eat the orange for her breakfast"). They were then presented with an image of a partly complete event outcome (i.e., a partly peeled orange) and asked if an event occurred (i.e., "Did she peel the orange?"). We manipulated the stated goal of the agent (High Goal, Low Goal, Neutral Goal) such that this goal required different degrees of physical completion of the sub-event in the test question to be satisfied.

We found that participants were more likely to accept a Partly Complete outcome as culminated if the outcome satisfied the agent's goal. Specifically, participants were more likely to treat events accompanied by Low Goal contexts as culminated compared to events accompanied by Neutral contexts. Overall, these findings suggest that, at least when visual information (e.g., the state of the object affected by the event) is sufficiently ambiguous, higher-order (goal) information can affect the construal of when an event ends.

One might find the lack of difference between Neutral and High Goal contexts puzzling. However, Neutral Goal contexts were not expected to represent a halfway point between the Low and High Goal contexts; indeed, much prior literature has assumed that, even in the absence of a specific context, an event such as peel the orange would culminate at its natural endpoint (Filip, 2017; see also Hindy et al., 2012; Solomon et al., 2015, among others). In practice, this means that High Goal contexts should introduce goals that need total or near-total completion of the subevent. However, the High Goal contexts of Experiment 1, while designed to require a high degree of subevent completion, did not actually necessitate strict completion. We addressed this issue in Experiment 2.

Experiment 2

In Experiment 2, we replicated Experiment 1 with stricter High Goal Contexts. We hypothesized that these suitably modified contexts would impose a higher completion threshold compared to having no explicit context at all (Neutral Goal condition), and could therefore lead to finer differentiation between the Context conditions.

Participants

Forty-two native English speakers were recruited from Prolific, an online recruitment platform. Participants in the study were paid $.76 for the 7 minute study.

We ran a simulation-based power analysis using the Experiment 1 data to estimate power for each of the planned Context contrasts in Experiment 2 (Neutral Goal vs. High Goal and Neutral Goal vs. Low Goal) following the method outlined in Kumle, Võ, and Drashkow (2021). The analysis was run using R2power, a function found in the mixedpower package (Kumle, Võ, & Drashkow, 2018) in R (R Core Team, 2021). R2power uses a user specified model and pilot data to simulate datasets and analyze power for varying numbers of participants for mixed effects models. We ran simulations for a number of participants based on Experiment 1 (n=42, with n=14 for each of 3 stimulus presentation lists) plus four other participant totals (21, 63, 84, 105). The reported power for each of these totals was calculated by taking the average power of 500 successful runs of the model for that given participant total. Of the seven possible participant totals tested, 42 participants were the smallest sample to achieve power over 80% for the Neutral Goal vs. Low Goal contrast (97.6%). The power for the Neutral Goal vs. High Goal contrast for the 42-participant category was 5.2% (the range across simulations was 4.6%-8.2%) but because

of the modified instructions for High Goals in Experiment 2, reliance on prior data for this contrast should be treated with caution.

Stimuli

Visual stimuli. The visual stimuli were identical to Experiment 1.

Verbal stimuli. The verbal stimuli were identical to Experiment 1, but the High Goal contexts now had to meet the constraint that the subevent in the test question had to be entirely complete for the goal to be achieved. Only 5 of our original contexts met this criterion and were included in Experiment 2; the remaining 13 were replaced with revised contexts. For instance, in the orange-peeling scenario depicted in Figure 1, the new High Goal context was: "Jesse wants to eat the orange but is allergic to the skin." See Section B of Supplemental Materials for a full list.

Procedure

The procedure was identical to Experiment 1.

Results

Visual Outcomes. The Visual Outcome data for Experiment 2 consisted of 42 participants x 54 items = 2,268 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 4 summarizes the data. The best fit for these data was a model that included Visual Outcome (Incomplete, Partly Complete, Complete) as a first level predictor. Table 4 presents the odds ratios for the multi-level model of Visual Outcome. For this analysis, Partly Complete outcomes were used as the comparison level. As expected, Partly Complete outcomes (M = 0.41) elicited

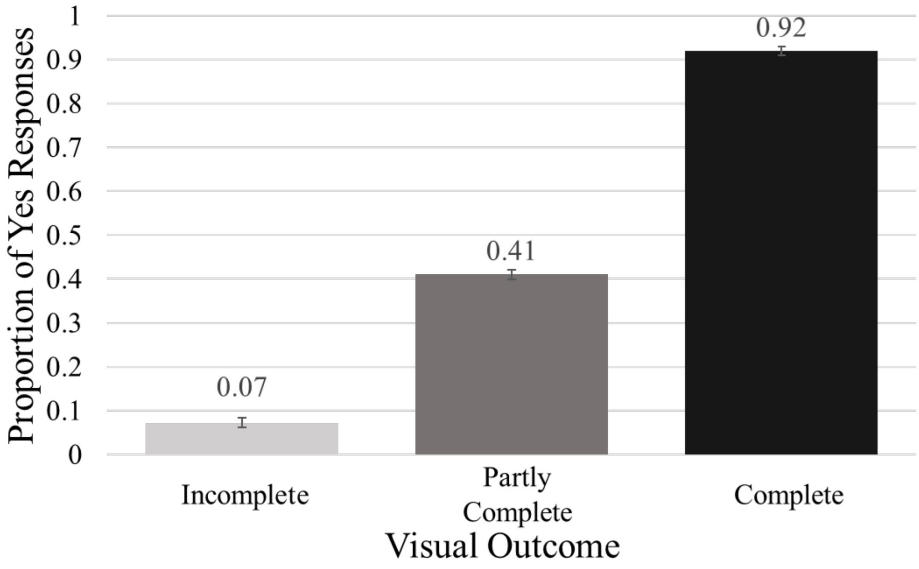


Figure 4. Proportion of Yes responses by Visual Outcomes (all items) in Experiment 2.

Table 4. Odds ratios for the multi-level model of Yes responses by Visual Outcomes (all items) in Experiment 2. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	0.63	0.18	-1.58
Incomplete	0.08	0.03	-6.20***
Partly Complete	1.00		
Complete	30.28	12.53	8.24***

Yes responses significantly more often than Incomplete outcomes ($M = 0.07$, $p < .001$) and significantly less often than Complete outcomes ($M = 0.92$, $p < .001$). As before, visual information affected the construal of event endpoints.

Linguistic Context. The Linguistic Context data for Experiment 2 consisted of 42 participants x 18 items = 756 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 5

summarizes the data. The best fit for these data was a model that included Context (Low Goal, Neutral Goal, High Goal) as a first level predictor. Table 5 presents the odds ratios for the multi-level model of Context. For this analysis, Neutral Goal was used as the comparison level. Low Goal contexts elicited Yes responses significantly more often ($M = 0.55$) than Neutral Goal contexts ($M = 0.40$, $p < .001$), and Neutral Goal contexts elicited Yes responses significantly more often than High Goal contexts ($M = 0.28$, $p < .001$).

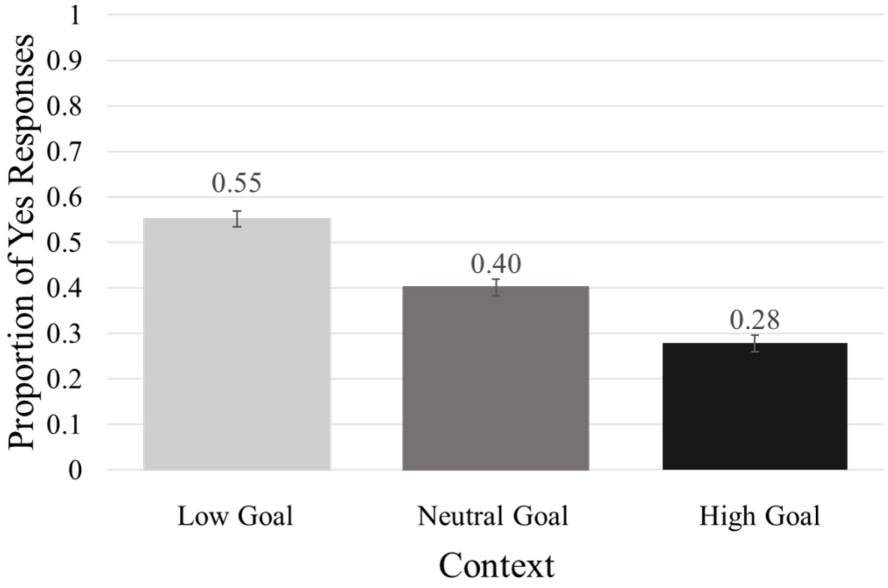


Figure 5. Proportion of Yes responses by Context (Partly Complete Outcomes only) in Experiment 2.

Table 5. Odds ratios for the multi-level model of Yes responses by Context (Partly Complete responses only) in Experiment 2. Significance levels: * p < .05, ** p < .01, *** p < .001

Effect	Odds Ratio	SE	z value
(Intercept)	0.58	0.22	-1.45
Low Goal	2.45	0.54	4.07***
Neutral Goal	1.00		
High Goal	0.45	0.10	-3.52***

Discussion

In Experiment 2, we modified the High Goal contexts used in Experiment 1 so that the subevent in the test question had to be entirely complete the agent's goal to be achieved. We found that participants were more likely to avoid treating a Partly Complete outcome as culminated if the outcome did not satisfy the agent's goal, and inversely, were more likely to accept the outcome as culminated if the outcome did satisfy the agent's goal. Overall, these findings further support the conclusion that higher-order (goal) information affects the perceived position of an event endpoint, at least when visual information (here, the degree of change in the affected object) is sufficiently ambiguous.

Experiment 3

Experiments 1 and 2 demonstrated that an agent's goals affected the mental placement of event endpoints for events that were not complete on visual grounds (judging by the change in the object involved in the event). We now explore whether this context effect generalizes to events where the visual cues supporting event completion are stronger (i.e., when the object undergoing change within an event is almost completely transformed). One possibility is that the context effects found in the previous experiments will be replicated for mostly complete events. Alternatively, the higher visual degree of event completion (or object-state change) might lead to the overall attenuation of the context effect. In Experiment 3, we tested these predictions by replicating Experiment 2 but replacing the partly complete visual outcomes (e.g., the partly peeled orange) with mostly complete ones (e.g., a mostly peeled orange).

Participants

Forty-two native English speakers were recruited from Prolific, an online recruitment platform. The sample size was the same as in Experiment 2. Participants in the study were paid $.76 for the 7 minute study.

Stimuli

Visual stimuli. The visual stimuli were identical to Experiment 2, but the Partly Complete target images were replaced with Mostly Complete target images for the same events as defined by the norming studies of Experiment 1. An example is given in Figure 6. Table 6 shows norming information for the images in Experiment 3.

Verbal stimuli. The goal contexts were identical to Experiment 2.

Procedure

The procedure was identical to Experiment 2.

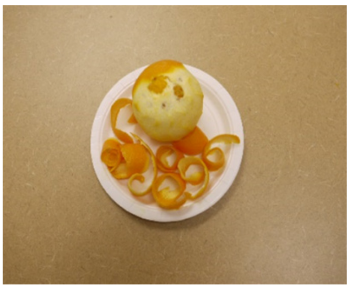


Figure 6. Example target (Mostly Complete) image for Experiment 3.

Table 6. Mean norming scores (and standard deviations) for Visual Outcomes included in Experiment 3. Scores for Incomplete and Complete stimuli are repeated from Table 1.

Visual Outcome	'What percent…?'	SD

Incomplete	7.91%	4.76%
Mostly Complete	69.84%	7.1%
Complete	92.78%	5.43%

Results

Visual Outcomes. The Visual Outcome data for Experiment 3 consisted of 42 participants x 54 items = 2,268 observations. We again used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 7 summarizes the data. The best fit for these data was a model that included Visual Outcome (Incomplete, Mostly Complete, Complete) as a first level predictor. Table 7 presents the odds ratios for the multi-level model of Visual Outcome. Mostly Complete outcomes were used as the comparison level. Mostly Complete outcomes ($M = 0.76$) elicited Yes responses significantly more often than Incomplete outcomes ($M = 0.07$, $p < .001$) and significantly less often than Complete outcomes ($M = 0.91$, $p < .01$). Again, we replicate the finding that visual information affected construals of when an event ends.

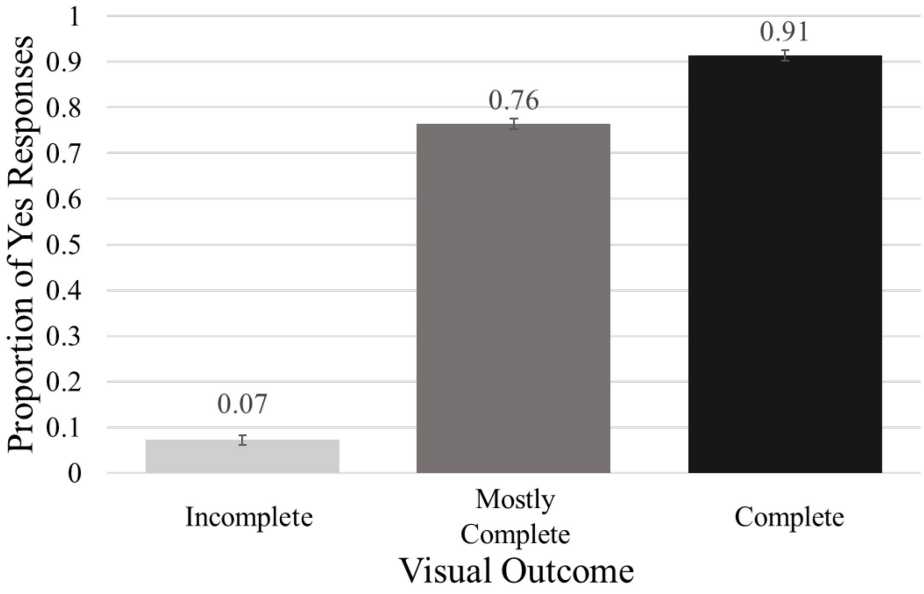


Figure 7. Proportion of Yes responses by Visual Outcomes (all items) in Experiment 3.

Table 7. Odds ratios for the multi-level model of Yes responses by Visual Outcomes (all items) in Experiment 3. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	4.31	1.04	6.07***
Incomplete	0.01	0.00	-13.84***
Mostly Complete	1.00		
Complete	3.98	1.28	4.31***

Linguistic Context. We used the subset of responses to Mostly Complete items in a separate analysis to look for effects of Context. These data consisted of 42 participants x 18 items = 756 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. Figure 8 summarizes the data. The best fit for these data was a model that included Context (Low Goal, Neutral Goal, High Goal) as a first level predictor. Table 8 presents the odds ratios for the multi-level model of

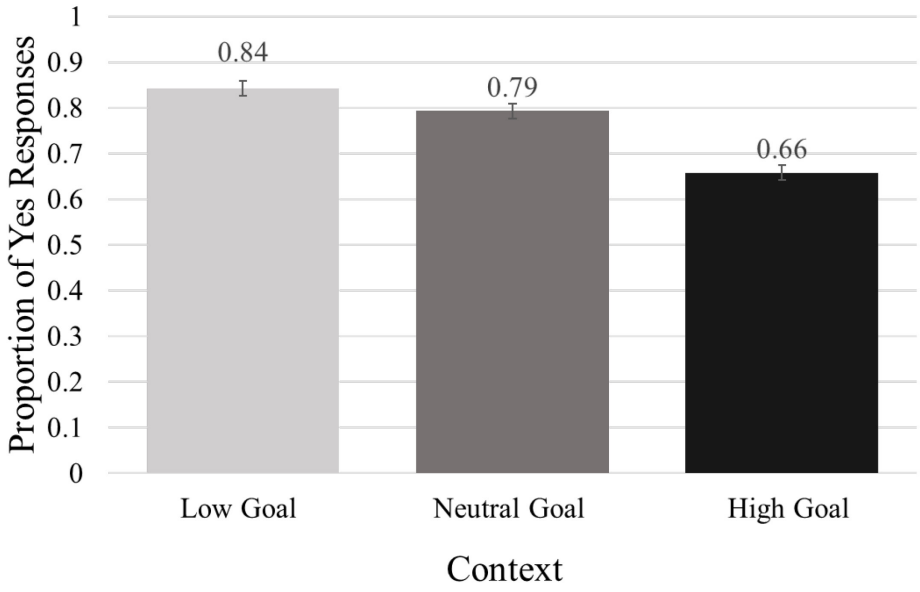


Figure 8. Proportion of Yes responses by Context (Mostly Complete Outcomes only) in Experiment 3.

Table 8. Odds ratios for the multi-level model of Yes responses by Context (Mostly Complete responses only) in Experiment 3. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	5.90	1.93	5.43***
Low Goal	1.40	0.38	1.24
Neutral Goal	1.00		
High Goal	0.39	0.09	-3.93***

Context. For this analysis, Neutral Goal was used as the comparison level. High Goal contexts ($M = 0.66$) elicited Yes responses significantly more often than Neutral Goal contexts ($M = 0.79$, $p < .001$). The difference in Yes responses between Low Goal contexts and Neutral Goal contexts was not significant ($p > .05$).

Comparison of Experiments 2 and 3. In a final analysis, we compared responses to target items across Experiments 2 and 3 (i.e., Partly and Mostly Complete items respectively) to see whether the physical degree of completion would affect responses to test questions in addition to the main factor of interest (Context). These data consisted of 84 participants x 18 items = 1,512 observations. We used a model that included Responses as the binary dependent variable and participants and items as crossed random intercepts. The best fit for these data was a model that included Context (Low Goal, Neutral Goal, High Goal) and Visual Outcome (Partly Complete, Mostly Complete) as first level predictors. The inclusion of an interaction term did not improve the model fit and was not included in the final model. Table 9 presents the odds ratios for the multi-level model of Context and Visual Outcome. In addition to the role of intentionality, these results confirm the expectation that the closer to its natural endpoint an event is, the more likely it is to be considered culminated by viewers, other things being equal.

Table 9. Odds ratios for the multi-level model of Yes responses by Context and Visual Outcome (target items only) across Experiments 2 and 3. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$

Effect	Odds Ratio	SE	z value
(Intercept)	0.64	0.21	-1.32
Context			
Low Goal	1.86	0.31	3.72***
Neutral Goal	1.00		
High Goal	0.45	0.07	-4.98***
Visual Outcome			
Partly Complete (Exp.2)	1.00		
Mostly Complete (Exp.3)	8.60	1.82	10.17***

Discussion

In Experiment 3, we found that participants were more likely to reject a Mostly Complete outcome if it failed to satisfy the agent's goal (see High Goal contexts). These findings support the conclusion that a salient goal can affect the mental placement of an event endpoint (Levine et al., 2017; Speer et al., 2004; 2007). They further suggest that goal-driven shifts in the mental placement of event endpoints do not apply only to highly ambiguous (or Partly Complete) visual stimuli (as in Experiments 1 and 2) but extend even to almost fully experienced events (Experiment 3).

A notable difference from the previous experiments is that now the Neutral Goal context patterned with the Low Goal context. This is reasonable given that the visual progression of event in the target items was such that it justified predominantly Yes responses even in the absence of a specific context (or when the context placed a minimal threshold for event completion). Indeed, a comparison of Experiments 2 and 3 found that the overall pattern of responses across Context conditions was similar regardless of the visual degree of completion, though Mostly Complete items were understandably more likely to be considered culminated overall.

General Discussion

Events make up every component of our daily lives, from making a cup of coffee in the morning to getting ready for bed at night. It has long been accepted that one cue that helps us to recognize when one event ends and another begins is the knowledge that an agent's goal has changed (Zacks & Tversky, 2001). Nevertheless, the role of higher-order goal information on event representations has remained poorly understood from the perspective of both event

segmentation and event conceptualization. Here, using a novel paradigm across a series of experiments, we asked whether prior knowledge of an agent's goals combines with visual cues to affect viewers' mental construal of an event endpoint (as assessed by viewers' answers to questions such as "Did she do X?").

We found that goal information affected endpoint construals for partly complete event outcomes, particularly for accepting an event as complete when the outcome satisfies an agent's goal (Experiment 1 and 2). Furthermore, we found that endpoints for visually mostly complete events are also subject to intentionality factors (Experiment 3). Overall, these findings offer the first direct piece of evidence in support of the conclusion that higher-order goal information affects how viewers conceptualize event endpoints.

Implications for Theories of Event Cognition

These findings have several implications for theories of event cognition. Within the context of past work that had recognized that both physical (Speer, Zacks, & Reynolds, 2004; 2007; Zacks & Tversky, 2001) and conceptual-intentional factors (Levine et al., 2017; Speer et al., 2004; 2007; Newtson, 1973; Vallacher & Wegner, 1987; Wilder, 1978) affect the placement of event boundaries, the present work is unique in successfully isolating the effects of higher-order (intentional) cues to events and their conceptual endpoints. Even though manipulating these kinds of cues individually is, of course, artificial (since, e.g., in real life, goals are embedded into specific situations), this approach is nevertheless important for understanding the foundations of event cognition. Our results support the conclusion that event endpoints are determined by a variety of considerations, some of which may be very abstract (cf. also Zacks & Tversky, 2001; Ji & Papafragou, 2020a, 2020b; Ünal, Ji & Papafragou, 2021).

Our results have more specific implications for models of event segmentation (e.g., Zacks et al., 2007). Recall that, despite the emphasis on how people identify event boundaries, these models do not discuss how people process the representational unit *within* event boundaries (i.e., "what happens" between the time an event begins and ends; see also Radvansky & Zacks, 2014; Elman & McRae, 2019; Cooper, 2021 for similar issues). For instance, EST (Zacks et al., 2007) speaks to the way in which people place event boundaries in a continuous stream of input, but this theory does not directly address event culmination. Our hypothesis is compatible with the basic insight from event segmentation theory, namely that event boundaries coincide with significant changes in event features (Swallow, Zacks, & Abrams, 2009; Zacks et al., 2007), but goes beyond this idea to show that the same degree of visual change in features of an event can be interpreted differently depending on intentionality factors. Thus our results throw light on the process of mentally assembling event units and their endpoints and the way this process yields an understanding of the conceptual content and structure of events. In particular, our data support an account in which tracking moment-by-moment perceptual and intentional changes is important for the construal and update of a single event representation and not simply for the replacement of one event representation with another one when an entirely new event begins (as traditionally claimed by EST; Zacks et al., 2007; cf. Kurby & Zacks, 2012). Our data provide evidence that a viewer's mental perspective on an event, and not simply the strictly observable characteristics of an event, contribute to how viewers construe events and make representational commitments about their endpoints – specifically, by shifting the mental placement of endpoints within the event timeline in accordance with intentionality cues.

The observed effects on the conceptualization of event endpoints bear on further claims of segmentation accounts. Event boundaries are known to facilitate the updating of event

information in working, long term, and procedural memory (Kurby & Zacks, 2008); furthermore, objects located at event boundaries have been shown to be remembered better than those located outside of an event boundary (Swallow et al., 2009; cf. also Strickland & Keil, 2011). Our findings raise the tantalizing possibility that intentionality-based cues to event endpoints might influence the way events are stored, processed and retrieved in memory: for instance, a visually incomplete event may be remembered as complete if the outcome satisfies the agent's intentions.

The present results also bear on IOH (Altmann & Ekves, 2019). As previously mentioned, IOH argues that an event must contain a change of state in some object across time and space (Hindy et al., 2012; Solomon et al., 2015). This change occurs independently of any observer. Our own results do show that the visual degree of event completion (and corresponding degree of change in the affected object, such as the peeled orange) influences intuitions about event endpoints. For instance, people's judgments in all our experiments are affected by visual cues about the development of an event (what we have called a *visual outcome* in our stimuli, indicating that the event has not been initiated, is partly underway or has terminated; see Visual Outcome analyses in all three experiments, and especially the analysis of the contribution of visual and intentional cues to event completion across Experiments 2 and 3). However, in a departure from the IOH view, our findings demonstrate that event representation is not uniquely tied to physical, observer-independent changes in an object but also depends on conceptual information such as an agent's goals.

Furthermore, the role of intentionality is not limited to situations in which object-based cues to event completion are ambiguous (a partly peeled object; Experiments 1 and 2): even for drastic and unambiguous object changes (e.g., a mostly peeled orange, Experiment 3) that should track the natural progression of the event timeline towards its natural endpoint, viewers'

intuitions about whether the event has reached its endpoint depend not on the visual, 'objective' degree of change in the object but on whether that change satisfies the agent's goals. As mentioned already, Altmann and Ekves (2019) allow for a role for intentionality and other abstract factors in event representation; however, this role is limited to situations in which an observer anticipates likely outcomes and states of objects based on perceived goals and intentions. Our findings clearly show that event cognition goes beyond representing objects and their affordances to engage social cognition: beyond perceptual features (e.g., visual object change), the conceptualization of event endpoints crucially integrates conceptual-intentional cues. [2]

How should we combine the contribution of physical, object-based and goal-related changes into future models of event representation? One possibility inspired by formal theories of object change and affectedness (Beavers, 2011; cf. also Hay, Kennedy & Levin, 1999; Wechsler, 2005; Kennedy & Levin, 2008; Hovav, 2008; Filip, 2017) would be to define change within an event as a transition of an event participant along a scale that defines the change. This 'change scale' can coincide with the degree of change in the event participant in a fairly direct way: in the natural course of things, the progress of peeling an orange coincides with changes in the object's state, culminating in a fully peeled orange. However, as we saw, the 'change scale' may not actually track the physical state of the object in the world that is affected by the event but something else that relates indirectly to it – such as the fulfillment of someone's intention

[2] Although not the main focus here, the present study also connects to discussions of event culmination in the psycholinguistic literature. When asked whether someone colored a picture, people are often likely to say yes even when the coloring is not complete (Arunachalam & Kothari, 2011; van Hout, 2018). This phenomenon has been observed in different languages (Jeschull, 2007; Li & Bowerman, 1998; Schulz & Penner, 2002; Weist, Wysocka, & Lyytinen, 1991) and characterizes both adults' and children's responses (van Hout, 2018; Jeschull, 2007; Schulz & Penner, 2002), but its origins are poorly understood. Our results suggest that the explanation for this phenomenon lies beyond narrowly linguistic (grammatical) factors and is part of how people reason about events and their boundaries given what they know about the goals of the agent in the event.

(goal): the progress of peeling an orange in that case is the degree to which the agent's goal is satisfied by different successive states of the orange. By introducing a more abstract notion of measurable scalar change into models of event cognition, it is possible to explain how event representations are incrementally updated in a flexible way that relies on both visual and non-visual (intentional) information as a dynamic stimulus unfolds. It can also explain how the timeline of even a seemingly observable event such as peeling an orange can depend on conceptual cues and can be construed as ending at different timepoints depending on context.

Limitations and Future Directions

Our study follows a long tradition of using a linguistic question to probe an event representation built on the basis of a prior visual stimulus (e.g., Hafri et al., 2013; cf. Griffin & Bock, 2000). The idea in our own and many other studies is that viewers extract information from the visual signal (alongside other cues) that allows them to later answer a linguistic test question about the event accurately. However, the current paradigm was not designed to assess the time course along which determinations of event culmination are made. As a result, this paradigm leaves open whether event culmination is computed prior to the presentation of the test question or whether the presence of the question influences participants' inferences. Similarly, we do not know whether our data reflect in-the-moment interpretations of event culmination or processes that involve event memory. Future work would need to replicate these findings with less overt and/or non-linguistic measures (e.g., reaction times or eye-tracking) to probe the time course of event endpoint construals. Relatedly, future versions of this work could ensure that these findings generalize to event representations that emerge as people process dynamically unfolding (and not only static) stimuli.

Our study relies on the ability to integrate linguistic and visual information about event structure and naturally connects to studies of events in verbal narratives (Zwaan et al., 1995; Zwaan & Radvansky, 1998; Bower & Rinck, 1999; Suh & Trabasso, 1993; Magliano & Radvansky, 2001; Trabasso & Wiley, 2005). Our findings further speak to a growing literature on how readers understand ongoing events in relation to characters' goals by integrating information within sequential visual narratives (i.e., comics, or picture stories; Cohn, 2020; Cohn & Paczynski, 2019; Kopatich et al., 2019), or within multimodal sequential narratives with both linguistic and visual components (Cohn, 2016; Cohn & Magliano, 2020). Adapting our current paradigm to further explore these issues would be an interesting next step in the effort to tease apart the individual contributions of conceptual and perceptual cues to event representation.

Our approach leads to the testable expectation that people could draw goal-driven event completion inferences from visual stimuli in the absence of language. Such inferences seem likely to occur routinely if warranted by context. In our day to day life, we often have to decide whether a seemingly 'concrete' event (clean the attic, wash the dishes, make the bed, open the door) has been accomplished. Often these construals rely on a single glance at the outcome (the state of the attic, the dishes, etc.) but are based on more than just the visual evidence afforded by the scene. That is, depending on the threshold placed by one's standards/goals, the very same visual information can be interpreted as indicating that the task was completed or still unfinished: for instance, cleaning the attic has a different endpoint depending on whether the goal is to dispose of unwanted belongings or to make the house presentable before selling it. Similarly, a half-opened door can be considered open if the goal is to let air in, or not open, if the goal is to let someone carrying heavy groceries through.

Viewed most broadly, the interplay between higher-order (including intentional) and perceptual considerations seems relevant for human conceptualization beyond the event domain. In a classic study by Labov (1973), people were shown a range of visual objects in the domain of tableware: cups, bowls, vases, and so on. In one manipulation, participants were provided with goals or functions for the objects (e.g., "Imagine in each case that you saw someone with the object in his [or her] hand, stirring in sugar with a spoon and drinking coffee from it"; "Imagine that you came to dinner at someone's house and saw this object sitting on the table, filled with mashed potatoes".) Participants' judgments about which objects counted as, for example, cups and bowls changed consistently in the presence of different goals and interacted with the visual appearance of the stimuli. Across domains, then, the representation of both spatial entities (objects) and temporal entities (events) by the human mind goes beyond observable properties of physical stimuli to integrate unobservable, social information about human interactions.

Concluding Remarks

In a set of experiments, we found that knowledge of an agent's goal affected the conceptualization of the point at which an event ended. These results strongly suggest that higher-order goal information affects the construal of even mundane and 'concrete' everyday events. Our findings place strong constraints on theories of event cognition by highlighting the rich and varied informational sources used by the human mind to represent event units.

Funding: This work was supported by National Science Foundation grant 2041171.

Acknowledgments: Part of this research was based on A.M.'s Master's thesis conducted while both authors were at the University of Delaware. We thank the members of the Language and Cognition lab at the University of Pennsylvania for helpful comments.

References

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

Arunachalam, S., & Kothari, A. (2011). An experimental study of Hindi and English perfective interpretation. *Journal of South Asian Linguistics*, *4*(1), 27-42.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Beavers, J. (2011). On affectedness. *Natural Language & Linguistic Theory, 29*, 335-370. https://doi.org/10.1007/s11049-011-9124-6

Bower, G. H., & Rinck, M. (1999). Goals as generators of activation in narrative understanding. In S. R. Goldman, A. C. Graessrt, and P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 111-134). Lawrence Earlbaum Associates, Inc.

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want. *Psychological Science*, *20*(1), 85–91. https://doi.org/10.1111/j.1467-9280.2008.02246.x

Cohn, N. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, *146*, 304-323. https://doi.org/10.1016/j.cognition.2015.10.007

Cohn, N. (2020). Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension. *Topics in Cognitive Science, 12*, 352-386. https://doi.org/10.1111/tops.12421

Cohn, N., & Magliano, J. P. (2020). Editors' Introduction and Review: Visual Narrative

Research: An Emerging Field in Cognitive Science. *Topics in Cognitive Science, 12,* 197-223. https://doi.org/10.1111/tops.12473

Cohn, N., & Paczynski, M. (2019). The neurophysiology of event processing in language and visual events. In R. Truswell (Ed.), *The Oxford Handbook of Event Structure* (pp. 623-637). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199685318.013.26

Cooper, R. P. (2021). Action production and event perception as routine sequential behaviors. *Topics in Cognitive Science, 13*(1), 63-78. https://doi.org/10.1111/tops.12462

van Dijk, T. A., & Kintsch, W. (1983). Strategies in discourse comprehension. New York: Academic Press

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*(2), 252-291. https://doi.org/10.1037/rev0000133

Filip, H. (2017). The semantics of perfectivity. *Italian Journal of Linguistics*, *29*(1), 167-200. https://doi.org/ 10.26346/1120-2726-107.

Griffin, Z. M. & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279. doi:10.1111/1467-9280.00255

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. https://dx.doi.org/10.1037%2Fa0030045

Hay, J., Kennedy, C. & Levin, B. (1999). Scalar Structure Underlies Telicity in "Degree Achievements". In T. Mathews and D. Strolovitch (Eds.), *SALT IX* (127-144). CLC Publications, Ithaca.

Hindy, N. C., Altmann, G. T., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object

state-changes on event processing: do objects compete with themselves? *Journal of Neuroscience*, *32*(17), 5795-5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012

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. https://doi.org/10.1075/tilar.24.05hou

Hovav, M. R, (2008). Lexicalized meaning and the internal temporal structure of events. *Theoretical and crosslinguistic approaches to the semantics of aspect, 110*, 13-42. http://dx.doi.org/10.1075/la.110.03hov

Huff, M., Meitz, T. G., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1377. https://doi.org/10.1037/a0036780

Jeschull, L. (2007). The Pragmatics of Telicity and What Children Make of It. In A. Belikova (Ed.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)* (pp. 180–187). Somerville, MA: Cascadilla 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*, 1465-1479. https://doi.org/10.1080/23273798.2020.1797839

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition*, 48(3), 390-399.https://doi.org/10.3758/s13421-019-00977-7

Kennedy, C., & Levin, B. (2008). Measure of Change: The Adjectival Core of Degree Achievements. In L. McNally and C. Kennedy (Eds.), *Adjectives and Adverbs: Syntax, Semantics and Discourse* (pp.156-182). Oxford: Oxford University Press.

Kopatich, R. D., Feller, D. P., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research: Principles and Implications*, *4*, 1-15. https://doi.org/10.1186/s41235-019-0176-1

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2018). Mixedpower: a library for estimating simulation-based power for mixed models in R. https://doi.org/10.5281/zenodo.1341047

Kumle, L., Võ, M. L., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior research methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition, 40*(5), 812-826. https://doi.org/10.3758/s13421-011-0179-8

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. N. Bailey & R. W. Shuy (Eds.), *New ways of analyzing variation in English.* Washington: Georgetown University Press.

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*. https://doi: 10.1080/23273798.2021.1896013

Levine, D., Hirsh-Pasek, K., Pace, A., & Golinkoff, R. M. (2017). A goal bias in action: The boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 916–927. https://doi.org/10.1037/xlm0000364

Li, P., & Bowerman, M. (1998). The acquisition of lexical and grammatical aspect in Chinese. *First Language*, *18*(54), 311-350. https://doi.org/10.1177/014272379801805404

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149. https://doi.org/10.1111/j.1467-7687.2008.00741.x

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*(5), 663-672. https://doi.org/10.3758/BF03196106

Magliano, J., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition*, *19*(1–2), 264–282. https://doi.org/10.1080/13825585.2011.633159

Magliano, J. P., & Radvansky, G. A. (2001). Goal coordination in narrative comprehension. *Psychonomic Bulletin & Review, 8*(2), 372-376. https://doi.org/10.3758/BF03196175

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. Cognition, 214, 104744. https://doi.org/10.1016/j.cognition.2021.104744

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Radvansky, G. A. & Zacks, J. M. (2014). *Event Cognition*. Oxford: Oxford University Press. https://doi.org/ 10.1093/acprof:oso/9780199898138.001.0001

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), https://doi.org/10.1111/cogs.12708

Santin, M. van Hout, A. & Flecken, M. (2021) Event endings in memory and language, *Language, Cognition and Neuroscience, 36*(5), 625-648, https://doi.org/10.1080/23273798.2020.1868542

Schulz, P. & Penner, Z. (2002). How you can eat the apple and have it too: Evidence from the acquisition of telicity in German. In J. Costa & M.J. Freitas (Eds.), *Proceedings of the GALA 2001 Conference on Language Acquisition* (pp. 239–246).

Solomon, S. H., Hindy, N. C., Altmann, G. T., & Thompson-Schill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience*, *27*(12), 2324-2338. https://doi.org/ 10.1162/jocn_a_00866

Speer, N. K., Zacks, J. M., & Reynolds, J. R (2004). Perceiving narrated events. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*. Retrieved from https://escholarship.org/uc/item/8m44j7g7

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries: Research article. *Psychological Science*, *18*(5), 449–455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a

matter of seconds. *Cognition*, *121*(3), 409-415. https://doi.org/10.1016/j.cognition.2011.04.007

Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language, 32*(3), 279-300. https://doi.org/10.1006/jmla.1993.1015

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), 236–257. https://doi.org/10.1037/a0015631

Trabasso, T., & Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes, 39*(2-3), 129-164. https://doi.org/10.1080/0163853X.2005.9651677

Tversky, B., Zacks, J. M., & Hard, B. M. (2008). The structure of experience. In T. F. Shipley and J. M. Zacks (Eds.), *Understanding events* (pp. 436-464). Oxford: Oxford University Press. https://psycnet.apa.org/doi/10.1093/acprof:oso/9780195188370.003.0019

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science, 13,* 224-242. https://doi.org/10.1111/tops.12475

Ünal, E., & Papafragou, A. (2019). How children identify events from visual experience. *Language Learning and Development, 15*, 138-156. https://doi.org/10.1080/15475441.2018.1544075

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Wechsler, S. (2005). Weighing in on Scales: A Reply to Goldberg and Jackendoff. *Language*

81(2), 465-493. https://doi.org/10.1353/lan.2005.0106.

Weist, R. M., Wysocka, H., & Lyytinen, P. (1991). A cross-linguistic perspective on the development of temporal systems. *Journal of Child Language*, *18*(01), 67. https://doi.org/10.1017/S0305000900013301

Wilder, D. A. (1978). Predictability of behaviors, goals, and unit of perception. *Personality and Social Psychology Bulletin*, *4*(4), 604–607. https://doi.org/10.1177/014616727800400422

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. http://dx.doi.org/10.1016/S0010-0277(98)00058-4

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979-1008. https://doi.org/10.1207/s15516709cog2806_5

Zacks, J. M., Bailey, H. R., & Kurby, C. A. (2018). Global and incremental updating of event representations in discourse. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, Madison, WI, USA, July 25-28.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., … Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655. https://doi.org/10.1038/88486

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

Zacks, J. M., & 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

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating

structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123(2),* 162 –185. https://doi.org/10.1037/0033-2909.123.2.162

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. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x