Boundedness is Represented in Visual and Auditory Event Cognition

Bahar Tarakçı (bahar.tarakci@ozyegin.edu.tr)
Özyeğin University, Istanbul, Turkey

Ceren Barış (ceren.baris@ozu.edu.tr)
Özyeğin University, Istanbul, Turkey

Ercenur Ünal (ercenur.unal@ozyegin.edu.tr)
Özyeğin University, Istanbul, Turkey

Abstract

Viewers are sensitive to the distinction between visual events with an internal structure leading to a well-defined endpoint (bounded events) and events lacking this structure and a well-defined endpoint (unbounded events). Here, we asked whether boundedness could be represented in the auditory modality in a way similar to the visual modality. To investigate this question, we trained participants with visual and auditory events on bounded or unbounded event categories in a category identification task. Later, we tested whether they could abstract the internal temporal structure of events and extend the (un)boundedness category to new examples in the same modality. These findings suggest that the principles and constraints that apply to the basic units of human experience in the visual modality have their counterparts in the auditory modality.

Keywords: event perception; event structure; boundedness; visual events; auditory events

Introduction

In our daily lives, we must constantly make decisions about the events that unfold around us based on various perceptual information. Some of these decisions require us to infer the end of an event. For example, when someone is in the bathroom, and we are waiting for them to leave, if we hear them flushing the toilet, it would be easier to infer that they will be out soon. However, if we hear the sound of running water, it might be more challenging to understand when they will finish taking a shower and leave the bathroom. Humans can readily distinguish events with clear boundaries and a well-defined endpoint from those that lack this structure based on visual information (Ji & Papafragou, 2020a; 2020b). An unexplored question in this field is whether principles that apply to event representations in the visual modality, such as categorizing them into distinct categories based on their internal temporal structure, transfer to the auditory modality. Here, we investigate whether listeners can also extract the temporal structure of events based on auditory information.

Event Representation in Cognition

Although people experience the happenings around them in a continuous manner, they divide this continuous experience into discrete event units with a beginning and an end (Zacks & Tversky, 2001). According to Event Segmentation Theory (Zacks et al., 2007), people form event models based on sensory inputs and conventional understanding of events. This event model is used for making predictions about what will happen next. When there is substantial conceptual or perceptual change during a specific temporal window, prediction error increases, making it difficult to predict upcoming events. This increase in prediction error causes the event model to be updated, and the update in the event model is perceived as an event boundary.

Event boundaries are privileged in event comprehension and memory compared to the temporal window between the event boundaries (for an overview, see Radvansky & Zacks, 2017). The ability to segment events into meaningful units among adults, which entails detecting boundaries, predicts event memory even when controlled for general cognitive abilities such as executive function, working, and episodic memory (Sargent et al., 2013; Zacks et al., 2006). People are better at describing the event and ordering depictions of the event when they see depictions of three consecutive boundaries, or breakpoints, compared to when they see depictions of three non-breakpoints (Newtson & Engquist, 1976). Objects seen at event boundaries are recognized better than objects seen between boundaries after a 5-second delay (Swallow et al., 2009; 2011). In the case of written narratives that contain consecutive events, retrieval of prior information is slower and less accurate if there is an event boundary between the encoded information and the memory probe (Speer & Zacks, 2005). Additionally, readers are slower to process clauses representing event boundaries (Pettijohn & Radvansky, 2016; Speer & Zacks, 2005; Zacks et al., 2009). Neuroimaging studies show that people have heightened neural activity bilaterally in the posterior cortex as well as in the right frontal cortex when boundaries appear in videos and narratives of events during intentional and naive viewings of events (Zacks et al., 2001; Speer et al., 2007; Whitney et al., 2009). Finally, in the domain of motion events, children and adults remember and detect changes to the endpoint of motion better than the point from which the motion originates (Lakusta & Landau, 2005; 2012; Papafragou, 2010).

Boundedness in Event Cognition

Events can be divided into two categories based on their internal temporal structure. Bounded events have natural boundaries and a well-defined terminal point or endpoint marking the event boundary (Comrie, 1976). Their internal temporal structure progresses towards this terminal point or culmination. If a bounded event is sliced into finer temporal

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

segments, each segment will correspond to a different stage of development leading to the event endpoint (Ji & Papafragou, 2020a; 2020b). This internal structure of bounded events has been characterized as as non-homogeneous or heterogeneous (Hinrichs, 1985). Bounded events naturally terminate when the terminal point or resultant state is reached. When someone flushes the toilet, the sound of the flush beginning marks a clear start to the event and the sound of the water refilling the tank marks a clear endpoint. Similarly, take the example of a bounded event shown in Figure 1a ('a man cracking an egg'). In this event, the action of cracking an egg has a well-defined end state. It is predictable when the event will end based on how the affected object appears after the action.

By contrast, unbounded events lack natural boundaries and a well-defined endpoint (Comrie, 1976; Kearns, 2000). Therefore, they may continue indefinitely and eventually terminate at any arbitrary moment. When someone starts taking a shower, the water starts running and this event can last indefinitely until that person decides to turn it off. If an unbounded event is sliced into finer temporal segments, each segment would still resemble each other as well as the main event (Jackendoff, 1991; Ji & Papafragou, 2020a; 2020b). For this reason, the internal structure of unbounded events has been characterized as "homogeneous" (Hinrichs, 1985). Take the example of an unbounded event shown in Figure 1c ('whisking an egg'). In this event, the action of whisking does not have a well-defined end state. In other words, it is not predictable when the event will terminate based on how the object appears in the resultant state.

Recent work has shown that humans are sensitive to the distinction between bounded and unbounded events in cognition. For instance, during a category identification task, participants were able to extract the boundedness category from videos of both bounded and unbounded events and extend these categories to new examples (Ji & Papafragou, 2020a). They were able to categorize events as bounded or unbounded even when truncated such that endpoints of the bounded events were not presented. Another important aspect of this study was that participants watched videos of bounded and unbounded event pairs marked with different colored frames during training. Because both (un)boundedness categories were marked during training, they needed to consider both categories while responding. Under these conditions, they were able to identify the bounded category more accurately than the unbounded one.

The sensitivity to boundedness is also reflected in how viewers respond to disruptions presented at different temporal points during event perception. In line with the asymmetry between bounded and unbounded events, these disruptions influence the way participants understand the temporal sequencing of bounded events more than they do of unbounded events, possibly due to the predictability of bounded events (Ji & Papafragou, 2020b). When perceiving dynamic events, participants had lower detection accuracy and slower reaction times for disruptions closer to the endpoint of the bounded events compared to disruptions at the event midpoint (Ji & Papafragou, 2022). This effect diminished or disappeared completely for unbounded events due to two differing characteristics of bounded and unbounded events. The homogeneity of the unbounded events causes each temporal segment to be as equally predictable as the other. Therefore, viewers treat disruptions at event midpoints and endpoints similarly. However, each temporal segment is another new and meaningful stage towards the event endpoint in bounded events.

Event Cognition Across Different Modalities

Although empirical evidence on the temporal profile of events is heavily based on visual event cognition, humans also receive information about events from other perceptual modalities, such as audition. It is possible that the principles and constraints that apply to the basic units of human experience in the visual modality have counterparts in the auditory modality.

This is clearly seen in the domain of objects: a fundamental unit of interest in the visual modality that bears critical parallels to events (Papafragou & Ji, 2023). Two representational systems encode objects as discrete individual entities (i.e., object file; Kahneman et al., 1992) or interconnected entities belonging to a group (i.e., ensemble representations; Im & Halberda, 2013) in the human mind. Empirical evidence demonstrates that auditory objects can be represented by these two representational systems, much like visual objects, though representations should be built on different object properties (Piazza et al., 2013). Moreover, linguistic cues that would bias either ensemble representations or individual object files influence the construal of visual and auditory objects similarly (Ongchoco et al., 2023).

One of the most fundamental ontological distinctions in object representation that have a relation to the notion of (un)boundedness is the distinction between objects and substances (Prasada et al., 2002). Object construals have a regular and repetitive structure across time and space, making their properties largely dependent on this structure. This non-arbitrary nature of objects contrasts with substances construals that are accidental configuration of entities that lack a clear structure. Unlike object construals, substance construals exhibit more arbitrary characteristics (Papafragou & Ji, 2023). This classification is also reflected in the temporal profile of events: bounded events resembling objects with quantifiable and countable nature and unbounded events resembling substances lacking this nature (Papafragou & Ji, 2023). In visual modality, viewers are able to extend (un)boundedness categories they extracted from events to examples of objects and substances (Papafragou & Ji, 2023). Furthermore, this connection between event and object construals could be drawn spontaneously without categorical training. Based on these findings, one open question is whether the ability to categorize events based on their internal event structure extends to the auditory modality in ways similar to the parallelity between visual and auditory modality in object representations.

The Present Study

In the present study, we take up this open question about how extracting (un)boundedness as a category extends to a less-studied perceptual modality, audition. To do so, we used sound and video recordings of the same events in a category identification task. We compared categorization accuracy across participants who only watched the videos of events to those who only listened to the sounds of events.

Our first aim is to test whether the auditory information would be sufficient to distinguish between bounded and unbounded events. If so, participants should be able to abstract event categories equally accurately from visual and auditory events.

As a secondary aim, we also test whether the asymmetry between bounded and unbounded events persists across modalities. If the principles that apply to perception of the event structure are similar across modalities, any asymmetries that emerge in the visual modality should also emerge in the auditory modality.

Method

Participants

We recruited a sample of forty-five native Turkish speakers (n = 45, 39 females, M_{age} = 21.10 years, SD_{age} = 2.39, range = 18.13 - 33.15). Participants were undergraduate students at Özyeğin University and received course credit in return for their participation. Data from two additional participants were excluded because their native language was not Turkish. The sample size was determined based on previous studies investigating boundedness in event cognition (Ji & Papafragou, 2020a; 2020b).

Materials

There were two types of stimuli: (i) video and (ii) sound recordings of events. The video recordings of events consisted of an actor performing an action on an object in front of a plain background without the sound of the action (e.g., cracking an egg vs. whisking an egg). The sound recordings consisted of the sounds of the exact same events (see Figure 1). We recorded video and sound recordings of all events simultaneously using a camera to record the visual versions and a separate phone to record the sounds to ensure the highest quality in both modalities. Video and sound files were synchronized in Adobe Premiere Pro. Later, the sounds were turned off in video recordings for the visual events. Sound recordings were edited in Audacity to ensure they were identifiable for the auditory events.

Eight pairs of auditory and visual events were used to train and test participants on the (un)boundedness categories. Each event pair consisted of a bounded and an unbounded event (see Figure 1). The duration of the events was held constant within the pair (range: 5.05 s - 11.93 s, M = 7.91 s). In each pair, the events differed based on the nature of the performed action. Therefore, bounded events depicted an action changing the state of an affected object (e.g., cracking an egg) for these events to have a well-defined beginning, midpoint, and endpoint. On the other hand, their unbounded counterparts depicted a continuous action that did not change the state of an object (e.g., whisking an egg). Therefore, the ending of these events could not be determined either due to the repetitive nature of the action. The sound profile of events reflected the bounded-unbounded distinction. For bounded events, the change in profiles displayed a clear beginning, midpoint, and endpoint in line with the smaller temporal segments representing different stages of development toward the event endpoint (see Fig.1b). In contrast, the sound profiles of unbounded events showed repetition, which demonstrates the homogeneous nature of the unbounded events (see Fig.1d).

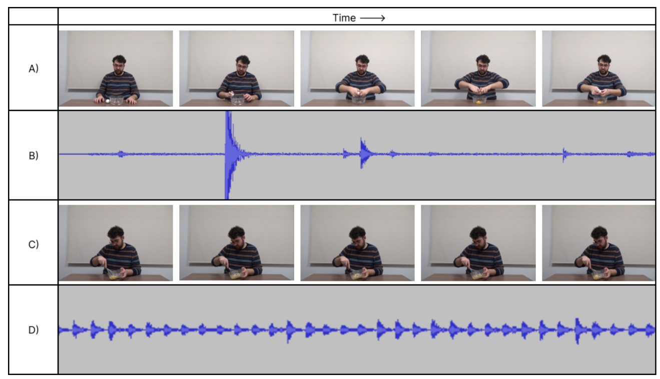

Figure 1: Visual and auditory stimuli of bounded and unbounded events: (A) screenshots from the beginning, midpoint, endpoint of video recording of breaking an egg (bounded); (B) sound profile of the sound recording of breaking an egg (bounded); (C) screenshots from the beginning, midpoint, endpoint of video recording of scrambling eggs (unbounded); (D) sound profile of the sound recording of scrambling eggs (unbounded).

Norming Study We conducted two norming studies to ensure all events were identifiable and reflected the bounded/unbounded distinction. None of the participants who took part in the norming studies participated in the main experiment.

The first norming study assessed whether participants could identify auditory and visual events. In this study, we obtained ratings from 21 native Turkish speakers for 18 event pairs with bounded vs. unbounded and visual vs. auditory versions. After watching/listening to the event recordings, participants were given a verb describing the action performed in the event and asked whether the video/sound they watched/listened to correspond to the indicated action. Based on the 0.5 exclusion criteria, we chose eight pairs of unbounded and bounded events with visual and auditory versions. In the final list, when participants indicated whether the verb they were presented with described the event they watched/listened to, there was no significant difference between bounded (M = 0.91, SD = 0.12) and unbounded (M = 0.95, SD = 0.10) events, $F(1, 28)$ = 1.28, p = .267. The difference between the visual (M = 0.96, SD = 0.09) and auditory (M = 0.89, SD = 0.12) modalities was not significant either, $F(1, 28)$ = 2.92, p = .099.

The second norming study assessed whether participants would view events as unbounded or bounded. In this study, we obtained ratings from 42 native Turkish speakers for 18 event pairs with bounded vs. unbounded and visual vs. auditory versions. After watching/listening to the event recordings, participants were asked whether the action they watched/listened to was an action with a beginning, middle, and end. At the end of the norming study, we excluded three unbounded events and their bounded counterparts because the average response to these events was higher than 0.5. In the final list with eight event pairs, when participants needed to indicate whether the event they watched/listened to had a beginning, a midpoint, and an endpoint, the bounded events (M = 0.91, SD = 0.08) elicited the "yes" response more than the unbounded events (M = 0.29, SD = 0.09), $F(1, 28)$ = 568.87, $p < .001$. There was no significant difference between auditory (M = 0.59, SD = 0.27) and visual events (M = 0.61, SD = 0.38), $F(1, 28)$ = 0.30, p = .591. However, there was an interaction between boundedness and modality, indicating that the difference between bounded and unbounded events was higher in visual modality than in auditory modality, $F(1, 28)$ = 15.12, $p < .001$ (see Table 1 for the final list of events).

Table 1: List of events used in the main study.

Phase	Bounded Events	Unbounded Events
Training	Opening a can of coke	Stirring coke with a straw
	Cracking an egg	Whisking an egg
	Pouring a cup of tea	Stirring tea
	Closing a metal box	Shaking a metal box
Testing	Eating soup with a spoon	Stirring soup with a spoon
	Drawing a square on a paper	Scribbling on paper
	Throwing a ball	Bouncing a ball
	Drinking orange juice	Shaking the orange juice bottle

Procedure

The main study was conducted face-to-face in a classroom. Participants watched or listened to the events together from a projector or a speaker and marked their responses individually on a response sheet. We manipulated two aspects of event categorization: (i) the modality and (ii) the (un)boundedness categories during training. Half of the participants watched videos, and the other half listened to audio recordings of the same events. The modality was consistent during training and testing for each participant. Within each modality group, we marked bounded events for half of the participants and unbounded events for the other half. The order of the events was pseudo-randomized. The location (top or bottom) of the bounded and unbounded events was counterbalanced within lists. The experimental procedure consisted of two phases: training and testing.

Training In the visual condition, participants watched four pairs of events consisting of a bounded and an unbounded event during the training. These pairs were matched based on the nature of the action. After watching both events one after the other, we told them that either the bounded or unbounded event would receive stars. This was accompanied by stars appearing around the video frame. The marked category remained consistent throughout the trials for each participant. In the auditory condition, participants listened to the sound recordings of events instead of watching their videos. Sound recordings were indicated using headphones and sound waves that signal which headphone the sound was coming from. In this condition, stars appeared around the headphones.

Testing In the visual condition, participants watched two bounded and two unbounded events during the testing. Further, each participant only saw one video of the bounded and unbounded event pair. Each video was presented by itself in the middle of the screen. After each event, we asked them whether the video would receive stars. In the auditory condition, participants listened to the sound recordings of events instead of watching their videos. In this condition, participants saw a single headphone in the middle of the screen and needed to indicate whether the auditory events they listened to would receive stars.

Results

All statistical analysis was performed in RStudio version 2023.12.0+369 (R Core Team, 2023). Data were analyzed with a generalized binomial linear mixed effects model using the *glmer* function of the *lme4* package (Bates et al., 2015).

The dependent variable was binary values for the accuracy of responses at the item level (1=accurate, 0=not accurate). The fixed effects included in the analysis were Boundedness (Bounded vs. Unbounded) and Modality (Visual vs. Auditory). The fixed effects of Boundedness and Modality were tested with sum-to-zero contrasts (-1/2, 1/2) (Schad et al., 2020). The model only included random intercepts for Items due to the singular fit error in a more complex model that included random intercepts for Subjects and Items. This error was resolved by excluding the random intercepts for Subjects so that the model would not be over-fitted and too complex for the data. Items were included as random intercepts instead of slopes because the between-subjects design of the study did not justify using random slopes.

This model revealed a significant intercept, indicating that overall participants were more likely to respond correctly than incorrectly (β = 1.315, SE = 0.222, z = 5.926, $p <$ .001). However, there were no significant differences in categorization accuracy for Bounded and Unbounded events

(β = -0.407, SE = 0.424, z = -0.960, p = .337) and across Auditory and Visual modalities (β = 0.160, SE = 0.423, z = 0.378, p = .706; Figure 2). There was also no significant interaction between Boundedness and Modality (β = 0.531, SE = 0.846, z = 0.628, p = .530). These results demonstrate that participants extracted bounded and unbounded categories equally accurately in visual and auditory modalities.

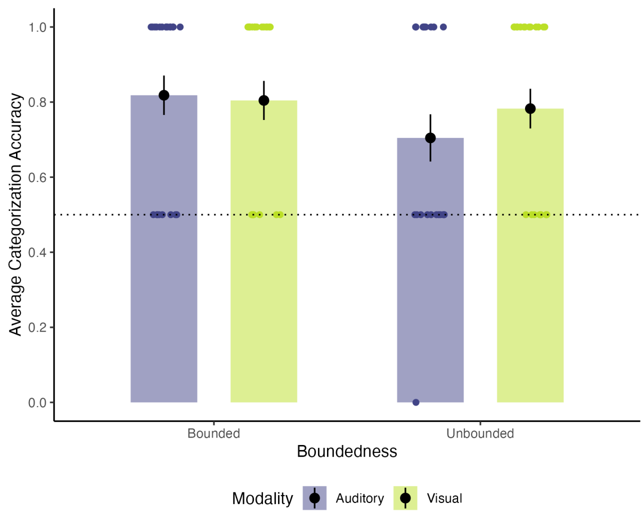

Figure 2: The height of the bars depicts average accuracy in test trials by boundedness and modality. Error bars indicate standard error of participant means. Colored dots represent the average accuracy for each participant.

Discussion

In this paper, we explored whether principles and constraints that apply to the basic units of human experience in the visual event representations have their counterparts in auditory event representations as they do in object representations. In relation to event representation, we asked whether auditory information would be sufficient to abstract the notion of (un)boundedness from several event examples. We further investigated whether the advantage of the bounded event category over the unbounded one would be present across visual and auditory modalities.

Regarding our first question, our participants were able to categorize bounded and unbounded events in both auditory and visual modalities equally well. When they saw or heard several examples of bounded and unbounded events, they were able to identify novel event examples belonging to the event category in which they were trained. These examples bore no resemblance except for their internal temporal structure. This means that they were able to interpret the abstract internal temporal pattern of events with only visual and auditory information. This finding extends the evidence in the spatial domain, indicating that auditory objects can be represented similarly to visual objects to another essential unit of human representation in the temporal domain, events (Ongchoco et al., 2023; Piazza et al., 2013). One possible explanation for the current findings could be that participants identified the event from their sounds and categorized them as bounded or unbounded based on their conventional understanding of the event structure. Another explanation for the similarity between visual and auditory events is that acoustic properties such as frequency, amplitude, and duration might play a role in understanding the event structure (Warren & Vebrugge, 1984). These properties of auditory events could be sufficient to abstract the temporal profile of events without visual access. Further research can benefit from using quantitative and qualitative approaches to decipher what qualities of auditory events facilitate how humans distinguish between bounded and unbounded events.

The similarity between the categorization of bounded and unbounded events across vision and audition highlights the importance of investigating event cognition in less-studied modalities. In specific cases, humans rely on auditory-only information as well as visual-only and audiovisual information about events in their daily lives. As previous research has heavily focused on visual or audiovisual events only, our findings are novel in showing that boundedness can be extracted from auditory-only information. These findings provide evidence for how events are perceived and categorized in cases where people rely solely on one modality when access to other modalities is limited for a specific amount of time. However, it also opens up the discussion on how people with a permanent lack of access to visual or auditory modalities throughout their lives could perceive and categorize events. For instance, in the case of congenital blindness, auditory information is the primary source of information for events and their spatio-temporal properties. Evidence from other domains suggests that sensitivity to auditory motion is preserved in a brain region related to visual motion processing in congenitally blind individuals as a type of cross-modal plasticity (Strnad et al., 2013). Our approach opens up possibilities for future work addressing such claims about cross-modal plasticity with sighted and congenitally blind individuals in the domain of event cognition.

Regarding our second research question, our participants were equally good at extracting bounded and unbounded categories across both modalities. Some studies in the past revealed that viewers were better at extracting the category of bounded events than the category of unbounded events when presented with visual events (Ji & Papafragou, 2020a). This asymmetry between bounded and unbounded event categories has been explained by cognitive and perceptual mechanisms that favor well-defined and discrete entities. Bounded events and objects are grouped because they consist of similar atomic features, such as discrete boundaries or minimal parts that can be quantified and counted (Papafragou & Ji, 2023). On the other hand, unbounded events are grouped with substances that consist of unspecified boundaries or atomic features. This difference between the nature of bounded and unbounded event categories has been taken into account as benefiting the identification and categorization of bounded events. However, there is evidence that the advantage of the bounded event category does not transfer to the categorization of objects vs. substances (Papafragou & Ji, 2023). If there are substantial parallels between the cognitive representations of objects and events, the lack of transfer from bounded events to objects raises questions about the asymmetry of (un)boundedness categories. In addition to this finding, one essential distinction between the current study and the studies that provide evidence for asymmetry is the number of categories they were trained with. In previous studies, both categories were marked with different colored frames during training. However, our study only marked one boundedness category at a time. Therefore, while the participants in the previous study needed to consider both categories during testing, our participants could focus on a single category. This contrast between the studies might have made it easier for our participants to extract the unbounded category, explaining why there is no asymmetry.

Finally, our findings connect to a broader discussion about how events are mapped onto language. The notion of boundedness not only characterizes how events are perceived and categorized but extends to how they are described, suggesting a homology between language and cognition in this domain (Ünal et al., 2021). In language, the distinction between bounded and unbounded events is systematically encoded through telicity (Filip, 2012; van Hout, 2016). For instance, a telic phrase such as 'making a cup of tea' describes an experience that has clear boundaries t and a structure with individuated sub-stages (e.g., boiling the water in the kettle, placing the teabag in a cup, pouring the boiled water into the cup, etc.). By contrast, an atelic phrase such as 'stirring the tea' describes a process and an experience that does not have clear boundaries. In fact, videos of bounded events elicit telic event descriptions, whereas videos of unbounded events elicit atelic event descriptions (Ji & Papafragou, 2020a). Although event descriptions in language map onto the distinction between bounded and unbounded event categories at the cognitive level, one thing to note is that these descriptions rely on visual information. One possible future direction for this line of research is exploring whether sensory and perceptual information from less-studied modalities could elicit event descriptions that systematically reflect bounded-unbounded distinction. The parallels between vision and audition in the temporal profile of events reported in this paper strongly suggest that the auditory modality would be a good test bed for investigating whether previously reported homologies between language and cognition extend to other perceptual modalities.

Acknowledgments

This work was supported by grant 121K259 awarded by TÜBİTAK to E. Ü. We thank Sezercan Uçar, Tuğçe Beyza Kirişçioğlu, Ecem Dokuyucu and Halil İbrahim Nebioğlu for their help with stimuli preparation and Lena Yazıcı for assistance in data entry.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

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

Filip, H. (2012). Lexical aspect. In R. I. Binnick (Ed.), *The handbook of tense and aspect.* Oxford University Press.

Hinrichs, E. W. (1985). *A compositional semantics for aktionsarten and np reference in english*. Doctoral dissertation, Department of Linguistics, The Ohio State University.

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278-286.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, *41*(1-3), 9-45.

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

Ji, Y., & Papafragou, A. (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience*, *35*(10), 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*, Article 104353.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175-219.

Kearns, K. (2000). *Semantics*. Palgrave Macmillan.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*(1), 1-33.

Lakusta, L., & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science*, *36*(3), 517-544.

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

Ongchoco, J. D. K., Yates, T. S., & Scholl, B. J. (2023). Event segmentation structures temporal experience: Simultaneous dilation and contraction in rhythmic reproductions. *Journal of Experimental Psychology: General*, *152*(11), 3266-3276.

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, *34*(6), 1064-1092.

Papafragou, A., & Ji, Y. (2023). Events and objects are similar cognitive entities. *Cognitive Psychology*, *143*, Article 101573.

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

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, *24*(8), 1389-1397.

Prasada, S., Ferenz, K., & Haskell, T. (2002). Conceiving of entities as objects and as stuff. *Cognition*, *83*(2), 141-165.

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133-140.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., ... Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241-255.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, Article 104038.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125-140.

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449-455.

Strnad, L., Peelen, M. V., Bedny, M., & Caramazza, A. (2013). Multivoxel pattern analysis reveals auditory motion information in mt+ of both congenitally blind and sighted individuals. *Plos One*, *8*(4), e63198.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, *23*(5), 1052-1064.

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.

Van Hout, A. (2016). Lexical and grammatical aspect. In J. Lidz, W. Synder, & J. Pater (Eds.), *The handbook of developmental linguistics.* Oxford University Press.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704-712.

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, *47*(1), 360-366.

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.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307-327.

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), 273-293. doi: 10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the alzheimer type. *Psychology and Aging*, *21*(3), 466-482.

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