Do Cross-Linguistic Differences Influence Event Perception?

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

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

Abstract

Telicity is an important semantic feature pointing to event construal: telic verb phrases denote bounded events with an inherent endpoint while atelic verb phrases denote unbounded events without such an endpoint. Languages encode telicity in different ways. Unlike English, Mandarin lacks an overt count-mass distinction and allows bare noun objects to form verb phrases. Would this cross-linguistic difference influence event perception? Experiment 1 elicited descriptions of bounded vs. unbounded events from English and Mandarin native speakers. A clear cross-linguistic difference was found: English speakers mostly used telic predicates for bounded events and atelic predicates for unbounded events while Mandarin speakers gave atelic predicates with bare noun objects for both event types. Experiment 2 explored how English and Mandarin speakers tracked the temporal structure of bounded vs. unbounded events. The two language groups performed similarly. The way people describe events may not affect the way they track event temporal profiles.

Keywords: telicity; cross-linguistic differences; universality; boundedness; event structure; event perception

Introduction

The world provides us with a dynamic and continuous flow of visual input but we perceive and describe this input in terms of structured units (i.e., events; Filip, 1993; Parsons, 1990; Zacks & Tversky, 2001). Research on event cognition has revealed that observers spontaneously segment continuous actions into events by attending to both perceptual cues (e.g., changes in location) and conceptual cues (e.g., causal relationships) (Magliano et al., 2001; Newtson et al., 1977; Zacks et al., 2007; Zacks & Swallow, 2007). Moreover, event boundaries, especially endpoints, are privileged over other time points in memory and language (for a review, see Radvansky & Zacks, 2017). For instance, viewers recognize objects relevant to an event boundary more accurately compared to objects relevant to non-boundary moments (Swallow et al., 2009). Similarly, when describing motion events, people include the goal (i.e., where a motion event ends) more often than the source (i.e., where a motion event begins; Do et al., 2020; Papafragou, 2010; Regier & Zheng, 2007; cf. Chen et al., 2023).

But what counts as an event endpoint? How do people process and represent different temporal slices of an unfolding event? Inspired by the linguistic literature (see later sections), a recent line of research uncovered an important distinction between *bounded events* that have an inherent, or natural endpoint (e.g., build a sandcastle) and *unbounded events* that lack such an endpoint (e.g., play with sand; Ji & Papafragou, 2020a). The temporal profile of bounded events can be divided into distinguishable stages. For example, building a sandcastle may include the steps of digging up sand, making the base, building walls, etc. By contrast, unbounded events have a largely undifferentiated internal structure. For example, playing with sand may involve scooping and pouring sand in a repetitive manner. Experimental evidence shows that viewers distinguish between these two event categories even when their use of language is blocked by a secondary linguistic task (Ji & Papafragou, 2020a). Furthermore, viewers track the temporal texture of events in event perception (Ji & Papafragou, 2020b, 2022). Specifically, endpoints are salient compared to other time points (such as the midpoints) in bounded but not unbounded events. In sum, boundedness (i.e., whether events have an inherent endpoint) is an important event property that shapes the way events are processed and represented (see also Filip, 1993; Folli & Harley, 2006; Malaia, 2014).

Research on boundedness and event cognition more broadly has so far been conducted almost exclusively with English speakers. However, event structure – including temporal structure - is encoded very differently across languages (e.g., Botne, 2003; Filip, 2004). To what extent could this variation affect event perception?

Temporal Structure of Events Cross-linguistically

In language, *telicity* refers to internal temporal event structure (Vendler, 1957; see Filip, 2012 for an overview). Telic predicates (*build a sandcastle*) describe events with different development stages evolving towards a "built-in terminal point" (Comrie, 1976) or "culmination" (Parsons, 1990). Such events have an inherent, or natural endpoint. By contrast, atelic predicates (*play with sand*) describe events with largely undifferentiated stages. Such events lack an inherent endpoint and can terminate arbitrarily.

In English, the distinction is shown in (1) and (2) below. An event of building a sandcastle comes to an end when a sandcastle came into being while an event of playing with sand does not specify how or when it ends. As the examples show, a telic predicate is congruent with a delimited temporal phrase (i.e., *in X time*) but incongruent with a durative temporal phrase (i.e., *for X time*), and an atelic predicate behaves in the opposite way (Dowty, 1979; Smith, 1991).

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.*

(1) a. A child built a sandcastle in half an hour. (telic)
b.*A child built a sand castle for half an hour.
(2) a. A child played with sand for half an hour. (atelic)
b.*A child played with sand in half an hour.

As shown in the examples, the telicity of a verb phrase depends compositionally on the nature of the action encoded in the lexical semantics of the verb as well as other verbal elements and the nature of the affected object(s) (for reviews, see Filip, 2012; Rothstein, 2004). First, telic predicates require verbs denoting an action leading to a change of state in the affected object, and thus the endpoint is naturally the resultant state. Atelic phrases involve verbs denoting actions that do not affect the object in a perceptible way. Second, telic predicates are formed by a quantified object (Krifka, 1989), as the changes in the object delimit or "measure out" the way the event develops (Tenny, 1987). In comparison, atelic phrases have direct objects unspecified about quantity.

In Mandarin, the telicity of a predicate can be tested through the possible locations of a temporal noun phrase (i.e., "*X time*", Duan, 2019; Ernst, 1987; Lin, 2008; Xiao & Mcenery, 2006). If a temporal phrase can appear before the verb but not before the direct object as in (3a-b), then it modifies a telic predicate and expresses similar meaning as the English temporal adverbial *in X time*. If a temporal phrase can appear before the direct object but not before the verb as in (4a-b), then it modifies an atelic predicate and behaves like the English temporal adverbial *for X time*.

(3) a. **ban-ge** **xiaoshi** jian-qi sha-bao.
half-CL hour build-rise sand-castle
"build a sandcastle/some sandcastles in half an hour"
b.*jian-qi **ban-ge** **xiaoshi** sha-bao.
build-rise half-CL hour sand-castle
*"build a sandcastle/some sandcastles for half an hour"

(4) a. wan **ban-ge** **xiaoshi** sha-zi.
play half-CL hour sand
"play with sand for half an hour"
b.***ban-ge** **xiaoshi** wan sha-zi.
half-CL hour play sand
*"play with sand in half an hour"

Similar to English, both the verb component and the noun object contribute to expressing telicity in Mandarin. First, a resultative verb compound (henceforth RVC) is widely used to explicitly encode the event end state and form a telic verb phrase. An RVC (e.g., *jian-qi* "build-rise" in 3a) is composed of two verbal elements, with the first one (e.g., *jian* "build") denoting an action and the second one denoting the result of the action (e.g., *qi* "rise", meaning "some construction rises", Li & Thompson, 1981; Sybesma, 1999). By contrast, mono-morphemic verbs (e.g., *wan* "play" in 4a) in Mandarin only express states or activities; they are inherently atelic (Lin, 2004; Sybesma, 1997; Tai, 1984; but see Smith, 1994; Soh & Kuo, 2005 for an alternative view).

Second, Mandarin lacks an overt count-mass distinction, and all nouns can appear in their bare form (Chierchia, 1998). Bare noun phrases are "indeterminate": depending on the context, they can be interpreted as referential (i.e., referring to entities identifiable in the context), or non-referential (i.e., not pointing to existent entities in the discourse) (Chen, 2014; cf. Soh & Kuo, 2005). Importantly, since a bare noun phrase does not carry any information about the quantity of its referent, it cannot delimit events. A bare noun (e.g., *sha-zi* "sand") and a mono-morphemic verb (e.g., *wan* "play") forms atelic verb phrases denoting an activity (e.g., *wan sha-zi* "play with sand" in 4a).[1]

In sum, to express whether an action leads to a salient change in state of the affected object, English speakers typically choose between different verbs, while Mandarin speakers choose between RVCs that specify the resultant state and mono-morphemic verbs that denote a process or an activity. Another major difference between English and Mandarin lies in expressing object quantity. When a single object is involved, English speakers have to use a quantified noun phrase since a bare singular form is ungrammatical (e.g., **build sandcastle*) but Mandarin speakers have the option of not specifying the quantity as bare nouns are allowed, as in (3a). If Mandarin speakers use bare noun phrases more often, they would give more atelic descriptions than English speakers. Specifically, the telic-atelic contrast in quantification of the affected object would be neutralized in Mandarin.

The Present Study

Could language-specific telicity patterns affect event perception? Recall that prior work on event cognition raises the possibility that boundedness is grounded in largely universal conceptual representations. On this view, language-specific patterns of encoding telicity would be unlikely to influence the perception of temporal event structure in non-linguistic tasks. An alternative possibility is that the conceptual representation of temporal event structure depends on linguistic encoding and should differ across speakers of different languages. According to this view, language-specific patterns of expressing telicity would have a global influence on the perception of event structure even when people are not explicitly using their native language. To test these two hypotheses, we compared how English and Mandarin speakers describe (Experiment 1) and perceive (Experiment 2) events with and without an inherent endpoint.

[1] A numeral and a classifier (e.g., *yi-ge shabao* "a sandcastle") can precede a noun to overtly specify the quantity of the referent in Mandarin. Numeral-classifier phrases form can delimit events and form a telic verb phrase (see further analysis and comparisons with other languages in Koenig & Chief 2008; Soh & Kuo 2005; Zhang, 2020, among others). The present study does not elaborate on number-classifier phrases for two main reasons. First, there is much similarity between such phrases and the quantified noun phrases in English in forming telic phrases. Second, such phrases in effect are not frequent in event descriptions; when there is no need to specify quantity, Mandarin speakers simply use bare nouns.

Experiment 1

Experiment 1 asked how English and Mandarin native speakers described bounded and unbounded events. Of interest was whether Mandarin speakers would be more inclined than English speakers to use atelic verb phrases with bare noun forms for both event types.

Method

Participants Thirty native speakers of English (age range: 18-23.5; M_{age}=19.2) and 30 native speakers of Mandarin (age range: 18-23; M_{age}=19.1) participated in the experiment. The English-speaking participants were undergraduates at a major university on the East Coast of the US. The Mandarin-speaking participants were students recruited from a major university in Beijing, China.

Stimuli We adopted the 20 pairs of videos showing bounded and unbounded events in Ji and Papafragou (2022; see Table 1). All videos involved a girl who did an everyday action which began with the girl picking up an object or tool from a desk and ended in her putting down the object or tool and moving her hands away from the table. Paired videos had the same duration (4.4-12.0s, M = 7.8s, SD = 2.4). Inspired by the linguistic literature, we created the contrast between bounded and unbounded events through two factors: the nature of the action and the nature of the affected object. For half of the videos, paired bounded and unbounded events involved the same object but differed in terms of the nature of the action the bounded event displayed an action that caused a clear change of state in the object (e.g., put up one's hair) while its unbounded counterpart did not involve such a change (e.g., scratch one's hair). For the other half of the videos, the bounded and unbounded events involved the same action but differed in terms of the nature of the affected object: the bounded event involved a single individual (e.g., draw a balloon) but its unbounded counterpart involved either an unspecified plurality of objects or a mass quantity (e.g., draw circles). Two norming studies showed that the bounded and the unbounded videos did not differ in the degree of intentionality or in the degree of visual similarity (Ji & Papafragou, 2020a). To ensure that people considered the stimuli as either a bounded or an unbounded event as designed, a third norming study eliciting judgment about the temporal structure of the stimuli was conducted. Videos of bounded events were considered as "something with a beginning, midpoint and specific endpoint" 87% of the time while videos of unbounded events were considered as such only 21.5% of the time (a significant difference, $t(39) = 20.33$, $p < .0001$; Ji & Papafragou, 2022).

The 20 pairs of events were split into 2 lists, such that each list included only one member of each pair with boundedness and the source of boundedness counterbalanced. In each list, bounded and unbounded events were intermixed such that items of the same event type could not appear successively more than 3 times. Two new lists were created where the order of the event items were reversed compared to the

Table 1: Event Stimuli in Experiment 1

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

original 2 lists. In sum, 4 event lists were used in the description task.

Procedure Participants were randomly assigned to one of the 4 event lists. They were instructed (in their native language) to watch each video and then to describe the video in a full sentence. Participants responded by writing down their sentences in an answer sheet.

Results

The descriptions of each language group underwent telicity diagnoses and were judged as telic or atelic by two RAs who were native speakers of English and of Mandarin, respectively. Overall, 95.7% of English descriptions and 92.3% of Mandarin descriptions received consistent telicity judgements. Discrepancies were then resolved by the two RAs and a third native speaker through discussion.

For bounded events, telic predicates were coded as target descriptions, all else as non-target descriptions. For unbounded events, atelic predicates were coded as target descriptions, all else as non-target descriptions. The binary data (whether a description was a target or a non-target one) were analyzed using logit mixed-effects models. Random intercepts were provided for each Subject and each Item (random slopes often did not converge, Baayen, et al., 2008; Barr, 2008; Barr et al., 2013). We examined the fixed effects of Language (English vs. Mandarin), Event Type (Bounded vs. Unbounded) and the Boundedness Source (Action vs. Affected object).

Results are shown in Figure 1. The two language groups performed differently, and overall English speakers gave more target descriptions (M=89.2%) than Mandarin speakers (M=59.8%) (β=-1.96, z=-8.46, p< .001). There were fewer target descriptions for bounded events (M=61.8%) compared to unbounded events (M=87.2%) (β =1.19, z=-6.63, p < .001).

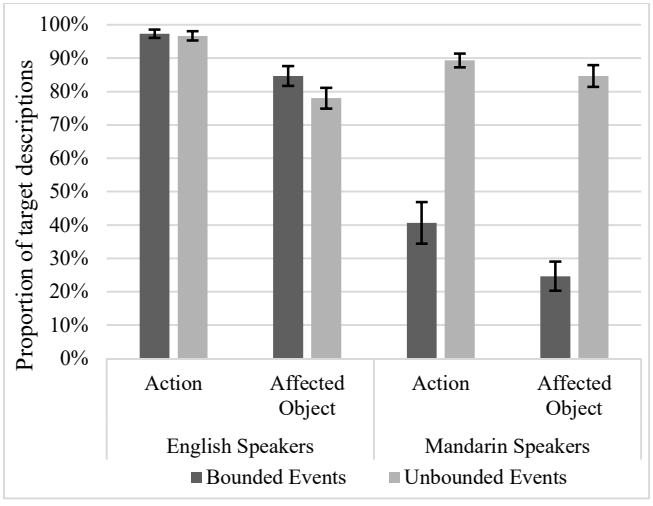

Figure 1: Proportion of target descriptions in Experiment 1. Error bars represent ±SEM.

A significant interaction between event type and language group was detected (β=3.22, z=8.92, p< .001): Mandarin speakers, but not English speakers had significantly fewer target descriptions for bounded events (Mandarin speakers: odds ratio=0.06, SE=0.01, p< .001; English speakers, odds ratio=1.52, SE=0.42, p= .132). As expected, English speakers mostly produced telic predicates for bounded events (M=91%), and atelic predicates for unbounded events (M=87.3%). By contrast, Mandarin speakers gave atelic descriptions for bounded events 54.3% of the time, and for unbounded events 87% of the time.

Comparing between the two sources of boundedness, there were more non-target descriptions for videos showing bounded-unbounded contrast in quantification of the affected objects (β=-1.35, z=-4.88, p< .001). A significant interaction between boundedness source and language group was detected (β=1.40, z=3.28, p= .001). Mandarin speakers, compared to English speakers, were less sensitive to the contrast in the quantification of the affected object; specifically, they used atelic phrases formed by mono-morphemic verbs and bare nouns for both bounded and unbounded events (odds ratio= 1.92, SE= 0.53, p= .017). For instance, *tie tie-zhi* ("stick sticker") was the most frequent (atelic) description for both the bounded event of sticking a sticker and its unbounded counterpart "sticking stickers".

Discussion

Results from English speakers reveal that the telic-atelic distinction aligns with the bounded-unbounded contrast shown in the videos. Specifically, stimuli of bounded events elicited telic descriptions and stimuli of unbounded events elicited atelic descriptions. Non-target descriptions were mostly telic predicates for unbounded events. Specifically, people sometimes specified the quantity of multiple objects involved in an unbounded event (e.g., "The girl cracked five peanuts."). Results from Mandarin speakers confirm the expectation that atelic phrases were used for describing both bounded and unbounded events. A major reason lies in the prevalence of bare nouns: for bounded events that involve a single object such as eating a pretzel, drawing a balloon, Mandarin speakers simply used atelic phrases such as *chi bing-gan* "eat pretzel", *hua qi-qiu* "draw balloon".

Experiment 2

Experiment 2 examined whether English and Mandarin native speakers tracked the temporal structure of events in similar ways. We adopted the break-detection task in Ji and Papafragou (2022). Very brief interruptions were inserted at either the temporal midpoints or close to the endpoints of videos showing bounded vs. unbounded events. Viewers were requested to detect these interruptions. Note that the interruptions were external to the event stimuli. Therefore, the detection performance would be lower when more processing resources were drawn by the event stimuli (see also Huff et al., 2012). Prior work with English speakers (Ji & Papafragou, 2022) has found that viewers exposed to bounded events were more likely to miss interruptions close

to the endpoints as their attention was drawn towards moments of event culmination while viewers of unbounded events detected midpoint and late interruptions equally well.

If boundedness is part of foundational and largely universal event representations, Mandarin speakers would perform similarly to English speakers given that the task of detecting interruptions was irrelevant to event perception and did not involve use of language. But if language-specific patterns of encoding telicity shape the way people compute boundedness, performance of the two language groups would differ. Specifically, Mandarin speakers' performance of detecting interruptions would be similar between the two event types.

Method

Participants Sixty-four native speakers of English (age range: 18-24; M_{age}=19.3) and 64 native speakers of Mandarin (age range: 18-23; M_{age}=19.2) participated in the experiment. The English-speaking participants were undergraduates at a major university on the East Coast of the US. The Mandarin-speaking participants were from a major university in Beijing, China.

Stimuli The video stimuli in Experiment 1 were edited in Corel VideoStudio to introduce a "break" of 0.03s (i.e., 1 editing frame, with a display rate of 30 frames per second. The break consisted of a blurry picture created by applying an Iris Blur Effect in Adobe Photoshop to portions of the original video (see the examples in Figures 2 and 3).

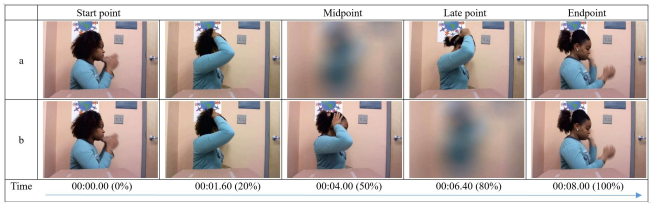

Figure 2: Examples of two versions of a bounded event (put up one's hair) in Experiment 2: (a) mid-break (b) late-break.

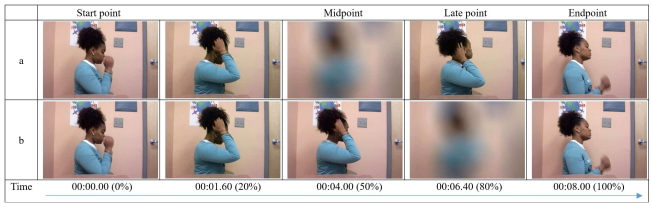

Figure 3: Examples of two versions of an unbounded event (scratch one's hair) in Experiment 2: (a) mid-break (b) late-break.

Each video was edited twice. In the mid-break version, the break replaced the frame of the temporal midpoint of the event (e.g., in the video of putting up one's hair with 240 frames, the mid-break replaced the 121st frame). In the late-break version, the break began at the point that corresponded to 80% of the event (e.g., in the same video of putting up one's hair, the late-break replaced the 193th frame). Edited videos with either a mid-break or a late break were used as test items, and their original versions were used as fillers.

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

Procedure Participants were randomly assigned to one of two conditions depending on the event type that they were exposed to throughout the experiment (Bounded or Unbounded). Within each condition, they were randomly assigned to one of the 4 lists. Participants were tested in groups of four to six in a lab room. They were told (in their native language) to watch each video carefully and decide whether they saw a break in the video. They responded by circling either "Break", or "No break" on an answer sheet. The text in the answer sheet was in their native language.

Participants were first given a practice session meant to illustrate what a break was. After each practice trial, participants noted their answer, and then the experimenter gave the correct answer. If participants were wrong, the video was played a second time. In the testing session, no feedback was given.

Results

We coded "Break" responses to test items and "No break" responses to fillers as correct. The binary accuracy data were analyzed using logit mixed-effects models. Random intercepts were provided for each Subject and each Item. All the two-level categorical predictors were coded with centering contrasts (-0,5, 0.5).

Performance on the filler items did not significantly differ between event types (Bounded events: M=94.1%; Unbounded events: M= 93.8%; β=-0.05, z=-0.19, p> .250), or between the two language groups (English speakers: M= 92.8%; Mandarin speakers: M= 95.1%; β=0.42, z=1.44, p= .151). Turning to test items, Language (English vs. Mandarin), Event Type (Bounded vs. Unbounded), and

Break Placement (Mid vs. Late) were included as predictors.[2] The interactions between the three variables were also examined and only the interactions that significantly improved the model fit were included in the final model.

As shown in Figure 4, the two language groups did not differ (English speakers: M=90.4%; Mandarin speakers: M=93.2%; β=0.37, z=1.39, p= .192). The difference between Bounded (M = 89.7%) and Unbounded event types (M = 94.0%) was also not significant (β=0.45, z=1.51, p= .130). Overall, participants were better at detecting mid-breaks (M=94.1%) than late-breaks (M=89.5%) (β=-0.53, z=2.06, p= .039). Importantly, a significant interaction between Event Type and Break Placement was detected (β= 1.78, z=3.44, p< .001): Participants watching videos of bounded events were better at detecting mid-breaks (M=95.3%) than late-breaks (M=84.0%) (odds ratio=4.16, SE=1.46, p< .001); by contrast, participants watching videos of unbounded events did not differ in their detection of mid-breaks (M=93.0%) and late-breaks (M=94.9%) (odds ratio= 0.70, SE=0.27, p > .250).

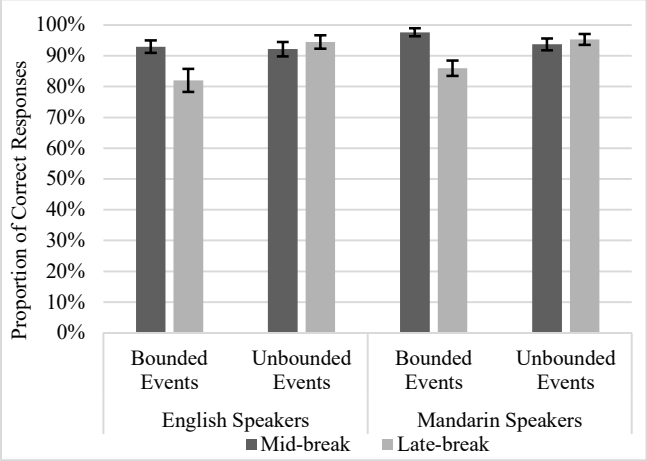

Figure 4: Proportion of correct responses in Experiment 2. Error bars represent ±SEM.

Discussion

No difference was found between English and Mandarin speakers' break-detection performance. In both language groups, people perceiving bounded events were more likely to miss a visual interruption when it occurred towards the end of an event; however, people exposed to unbounded events could detect interruptions at midpoints or close to endpoints equally accurately. Therefore, the results support the hypothesis that boundedness is a basic conceptual distinction that affects event perception in similar ways cross-linguistically.

General Discussion

Our findings show that English and Mandarin speakers differ in their use of telic vs. atelic phrases when describing events. Unlike English speakers, Mandarin speakers are inclined to use atelic phrases for both bounded and unbounded events as Mandarin allows bare noun forms. Despite the differences in event descriptions, both groups perceive bounded vs. unbounded events in similar ways: they track the internal temporal contour of unfolding events and process different temporal slices within an event accordingly. Taken together, these findings support the hypothesis that boundedness is a foundational property in conceptual representations of events, independent of language-specific patterns of encoding event internal temporal structure.

Our data suggest that boundedness may serve as a conceptual basis for the linguistic notion of telicity. They also raise interesting questions about how shared representations of (un)boundedness interface with the need to acquire and use different lexical and/or grammatical means to encode telicity in the present language samples – but also across languages (Bar-El et al., 2004; Botne, 2003; Filip, 2004; Kardos, 2016; Singh, 1998; Soh & Kuo, 2005; Zhang, 2020).

Finally, our findings contribute to efforts to disentangle language-specific from language-independent (potentially universal) components of event structure, and more generally connect language to event cognition (see also Flecken et al, 2015; Gerwien & von Stuterheim, 2018; Konishi et al., 2019; Sakarias & Flecken, 2019, among many others). We note that our event perception task was purposefully constructed to provide an implicit test of sensitivity to boundedness. In that sense, it offers a strong test of event perception that lends itself to testing a variety of populations cross-linguistically beyond the English-speaking participants in many past event cognition studies.

Our data connect to a recent study by Santin et al. (2021) that investigated how Spanish and Mandarin speakers talked about and remembered event endings. Their results show that culminated events (i.e., bounded events where the inherent endpoints were achieved) were remembered best compared to ongoing actions and events without a clear change of state (i.e., unbounded events) by both language groups. This finding is in line with our work, providing consistent evidence for the salience of endpoints in bounded events cross-linguistically. Santin and colleagues (2021) also detected a moderate effect of language-specific properties on memory of event culmination: the use of single verbs by Spanish speakers to lexicalize the culmination moment could enhance memory to a greater extent compared to the use of RVCs by Mandarin speakers. This language effect was detected in a widely-used paradigm where people describe events and then are tested on their event memory (see also Gennari et al., 2003; Skordos et al., 2020). The present results leave open the possibility that the overt use of telic vs. atelic phrases in our samples can affect subsequent event perception or memory.

[2] Adding List, Gender, Boundedness Source (Action vs. Affected Object), or any interaction between Boundedness Source and other predictors to the model did not reliably improve model fit so we excluded these factors from further analysis.

Acknowledgments

This material is based upon work supported by the National Social Science Fund of China grant #21CYY046 (Y.J.) and National Science Foundation Grant BCS-2041171 (A.P.).

References

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

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

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

Botne, R. (2003). To die across languages: Towards a typology of achievement verbs. *Linguistic Typology, 7*(2), 233–278.

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

Chen, P. (2014). Referentiality and definiteness in Chinese. In C.-T. Huang, Y.-H. Li, & A. Simpson (Eds.) *The handbook of Chinese linguistics*. Wiley Blackwell.

Chen, Y., Trueswell, J., & Papafragou, A. (2024). Sources and goals in memory and language: Fragility and robustness in event presentation. *Journal of Memory and Language, 135,* 104475.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics, 6*, 339–405.

Do, M. L., Papafragou, A., & Trueswell, J. (2022). Encoding motion events during language production: Effects of audience design and conceptual salience. *Cognitive Science, 46,* e13077.

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

Duan, W. (2019). *On telicity of eventualities in Mandarin Chinese.* [Unpublished doctoral dissertation]. Zhejiang University.

Ernst, T. (1987). Duration adverbials and Chinese phrase structure. *Journal of Chinese Language Teachers Association, 22,* 1–11.

Filip, H. (1993). *Aspect, situation types and nominal reference* [Unpublished doctoral dissertation]. University of California at Berkeley.

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

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

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

Folli, R., & Harley, H. (2006). What language says about the psychology of events. *Trends in Cognitive Science, 10,* 91–92.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*, 49–79.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition, 180*, 225–237.

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

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

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

Kardos, É. (2016). Telicity marking in Hungarian. *Glossa: A Journal of General Linguistics. 1*(1), 1–37.

Konishi, H., Brezack, N., & Golinkoff, R. M. (2019) Crossing to the other side: Language influences children's perception of event components. *Cognition, 190,* 104020.

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

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

Li, C. N., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Lin, J. (2004). *Event structure and the encoding of arguments: The syntax of the Mandarin and English verb phrase* [Unpublished doctoral dissertation]. MIT.

Lin, J.-W. (2008). Event decomposition and the syntax and semantics of durative phrases in Chinese. In J. Döling, T. Heyde-Zybatow, & M. Schäer (Eds.), *Event structures in linguistic form and interpretation*. Mouton deGruyter.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*(5)*,* 533–545.

Malaia, E. (2014). It still isn't over: event boundaries in language and perception. *Language and Linguistics Compass, 8*, 89–98.

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*(12)*,* 847–862.

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

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

Radvansky, G., & Zacks, J. (2014). *Event cognition*. OUP.

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

Rothstein, S. (2004). *Structuring events*. Malden: Blackwell Publishing.

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

Santin, M., van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience, 36*(5), 625–648.

Singh, M. (1998). On the semantics of the perfective aspect. *Natural Language Semantics, 6*(2), 171–199.

Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology, 37*, 254-270.

Smith, C. S. (1991). *The parameter of aspect*. Kluwer.

Smith, C. S. (1994). Aspectual viewpoint and situation type in Mandarin Chinese. *Journal of East Asian Linguistics, 3*, 107–146.

Soh, H. L., & Kuo, J. Y-C. (2005). Perfective aspect and accomplishment situations in Mandarin Chinese. In H. Verkuyl, H. Swart, & A. van Hout (Eds.), *Perspectives on aspect*. Springer.

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

Sybesma, R. (1997). Why Chinese verb LE is a resultative predicate. *Journal of East Asian Linguistics,* 6, 215–261.

Sybesma, R. (1999). *The Mandarin VP*. Kluwer.

Tai, J. H.-Y. (1984). Verbs and times in Chinese: Vendler's four categories. In D. Testen, V. Mishra, & J. Drogo (Eds.), *Papers from the parasession on lexical semantics*. Chicago Linguistics Society.

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

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

Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A corpus-based study*. John Benjamins Publishing Company.

Zacks, J. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 26*(6)*,* 979–1008.

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

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

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

Zhang, A. (2020). Referentiality, individuation and incompletive readings. *Journal of East Asian Linguistics, 29,* 435–468.