Events and Objects Are Similar Cognitive Entities

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

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

Abstract

Logico-semantic theories have long noted parallels between the linguistic representation of temporal entities (events) and spatial entities (objects): bounded (or telic) predicates such as *fix a car* resemble count nouns such as *sandcastle* because they are "atoms" with well-defined boundaries. By contrast, unbounded (or atelic) phrases such as *drive a car* resemble mass nouns such as *sand* in that they are unspecified for atomic features. Here, we show for the first time that there are similarities in the *perceptual-cognitive* representation of events and objects in non-linguistic tasks. Specifically, after viewers form a bounded or an unbounded event category, they can extend the category to objects or substances respectively (Experiment 1). Furthermore, viewers can intuitively make event-to-object mappings that respect atomicity (Experiment 2). These striking similarities between the mental representation of events and objects have implications for current theories of event cognition, as well as the relationship between language and thought.

Keywords: boundedness; event cognition; telicity; aspect; atomicity; object

Introduction

Humans encode their dynamic experience as concrete units, i.e., events. According to an influential account (Zacks, et al., 2007), segmenting events from continuous streams of actions depends on stable working memory representations, known as event models. Event models contain abstract information about events such as the number of event participants, their roles and intentions, their temporal and spatial relations (Radvansky & Zacks, 2014). Rich evidence from event segmentation studies shows that event boundaries (especially the event endpoints) are important for event perception and comprehension (e.g., they are recalled and described more accurately than non-boundaries; Huff et al., 2012; Newston & Engquist, 1976; Swallow et al., 2009; cf. also Lakusta & Landau, 2005, 2012; Levine et al., 2017; Papafragou, 2010; Regier & Zheng, 2007).

Although vast research has demonstrated how event observers form event units and has accounted for the importance of event boundaries, much less attention has been paid to the representational content of an event unit (i.e., "what happens" between two event boundaries). Event models (Radvansky & Zacks, 2014) and related frameworks (Cooper, 2021; Elman & McRae, 2019) have only addressed broad parameters of complex event knowledge. Other work has revealed how infants, children and adults conceptually represent dynamic events but has targeted specific event types and ingredients (e.g., motion – Lakusta & Landau, 2005, 2012; Papafragou, et al., 2008; Pulverman et al., 2013; spatial mechanics – Baillargeon & Wang, 2002; McDonough et al., 2003; Strickland & Scholl, 2015; causation – Buchsbaum et al., 2015; Muentener & Carey, 2010; Scholl & Nakayama, 2002; Wolfe, 2007; and intentionality – Tomasello & Carpenter, 2007; Woodward & Cannon, 2013; among many others). In sum, general models of event representation still need to specify how the process of mentally assembling event units and their boundaries yields an understanding of the conceptual content and organization of events.

A separate linguistic and philosophical tradition on events that goes back to Aristotle may offer insights on this research gap. Language describes a dynamic situation as either a bounded or an unbounded event through the telic-atelic distinction (see Filip, 2012; van Hout, 2016). For instance, the telic sentence *Sam entered the house* encodes an experience as a bounded event: this event has internal structure consisting of distinct stages (e.g., opening the door, taking a step in, etc.) that lead to a definite endpoint (Jackendoff, 1991; Mourelatos, 1978; Parsons, 1990; Vendler, 1957) - the moment when Sam is in the house. By contrast, the atelic sentence *Sam approached the house* encodes an experience as an unbounded event: in this event, "any part of the process is of the same nature as the whole" (Vendler, 1957) – each moment of the action can still be described as an event of approaching the house. As a result, unbounded events have no specified endpoint and may terminate at an arbitrary moment. Several studies have confirmed that telic vs. atelic language triggers different cognitive perspectives on the temporal structure of events (Barner et al., 2008; Malaia et al., 2012; Strickland et al., 2015; Wagner & Carey, 2003).

A cornerstone of analyses of linguistic telicity extends beyond the event domain to connect to the linguistic semantics of nominals (e.g., Bach, 1986; Taylor, 1977). On this view, bounded (or telic) predicates such as *enter (a house)* resemble count nouns such as *sandcastle* because they are "atoms" that can be individuated and compared to each other; these entities have a non-homogeneous structure (i.e., they have discrete minimal parts and cannot be divided arbitrarily). By contrast, unbounded (or atelic) phrases such as *approach (a house)* resemble mass nouns such as *sand* in that they are unspecified for atomic features; these entities are homogeneous (i.e., they lack minimal parts and can be divided arbitrarily). Atomic (individuated) entities (bounded events and objects) can be distributively quantified, counted, and compared by their number (e.g., to answer a question about bounded events such as "Who gave more kisses?", people directly count how many kisses were given; similarly, a question about objects such as "Who owns more books?" is solved by simply counting the books; Barner et al., 2008; Wittenberg & Levy, 2017). Non-atomic (unindividuated) entities (unbounded events and substances) cannot easily be distributively quantified, are better measured than counted, and are not preferentially compared by number (e.g., to answer "Who did more running?", people need measurements such as the distance or time duration of running but number becomes irrelevant; similarly, to answer "Who has more milk?", one may check the volume or weight of milk rather than consider number).

Evidence for the parallel between the cognitive representations of events and objects has so far mostly come from studies of how linguistic form connects to conceptualization (Barner et al., 2008; Prasada, et al., 2002; Wellwood et al., 2018). In the present study, we for the first time directly tested the hypothesis that the cognitive unit of event representation is similar to the unit of object representation. Specifically, we asked whether abstract considerations of atomicity underlie the domains of both temporally extended entities (bounded/unbounded events) and spatially extended entities (objects/substances) in non-linguistic cognition. In Experiment 1, we built on the boundedness studies by Ji and Papafragou (2020) to test whether people can extend the category of event (un)boundedness to a corresponding quantificational distinction in the object domain. In Experiment 2, we examined whether people can match events to objects based on the (non-)atomicity property through a more direct task.

Experiment 1

In this experiment, we conducted a category identification task (adapted from Ji & Papafragou, 2020), where participants watched pairs of bounded and unbounded events and had to form a generalization about one member of these pairs. Later participants were invited to extend their event-based generalizations to novel objects and substances. Of interest was whether participants would use the logic of individuation to generalize from bounded events to objects and from unbounded events to substances.

Method

Participants Forty-eight adults (34 female, 14 male; M_{age} = 19.5, age range: 18.0 – 23.2) participated in the experiment. All were undergraduates at a major research university on the East Coast of the United States and received course credit for participation.

Stimuli We used 10 pairs of videos from Ji and Papafragou (2020). All videos involved the same girl doing a familiar everyday action in a lab room. Paired videos had the same duration (4.4-12s, M = 7.3s), and showed a bounded and an unbounded event (see Table 1). The bounded-unbounded contrast was due to the nature of the action. Specifically, in each pair, the bounded member involved an action that caused a clear and temporally demarcated change of state in the object (e.g., dress a teddy bear) while its unbounded counterpart did not have such a change (e.g., pat a teddy bear; see Figure 1). Paired events included the same affected object(s) (e.g., a teddy bear). Among the 10 pairs of videos, 5 pairs involved a single canonical object (e.g., dress vs. pat a teddy bear) and 5 pairs involved a mass quantity (e.g., put up vs. scratch one's hair).

Table 1: Stimulus events in Experiment 1

Phase	No.	Bounded Events	Unbounded Events
Training	1	fold up a handkerchief	wave a handkerchief
	2	put up one's hair	scratch one's hair
	3	stack a deck of cards	shuffle a deck of cards
	4	group pawns based on color	mix pawns of two colors
Testing	5	dress a teddy bear/	pat a teddy bear
	6	roll up a towel/	twist a towel
	7	fill a glass with milk/	shake a bottle of milk
	8	scoop up yogurt/	stir yogurt
	9	close a fan/	use a fan for oneself
	10	crack an egg/	beat an egg

Note: Each row depicts a pair of events. In the training phase, participants saw both events within a pair. In the testing phase, participants saw only one event from each pair (as indicated by dashes). We rotated the videos used in the training (No. 1-4) and the first four pairs in testing phase (No. 5-8).

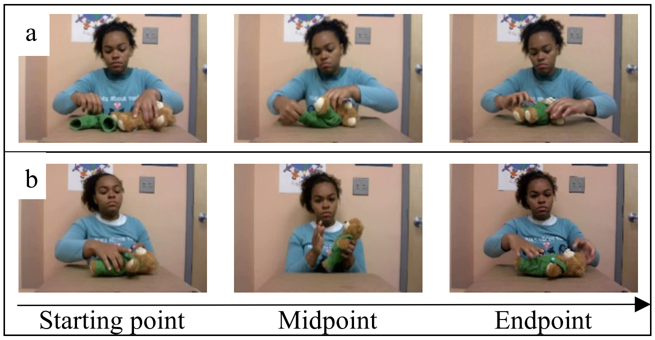

Figure 1: Examples of paired videos in Experiment 1: (a) a bounded event (dress a teddy bear), (b) an unbounded event (pat a teddy bear).

Two types of norming studies were used to confirm the placement of the stimuli within the bounded vs. the unbounded class. First, videos of bounded events were considered as "something with a beginning, midpoint and specific endpoint" 90.0% of the time while videos of unbounded events were considered as such only 17.5% of the time ($t(18) = 17.99, p < .001$). Second, in an event description task, videos of bounded events elicited telic descriptions 97.8% of the time, and stimuli of unbounded events elicited atelic verb phrases 95.6% of the time ($t(18) = 1.09, p > .250$); the former included change-of-state predicates (e.g., dress a teddy bear) and the latter verbs of activity (e.g., pat a teddy bear). An additional rating study invited participants to judge how intentional the action in each video looked on a scale from 1 (totally unintentional) to 7 (intentional). The results indicated that events within the bounded and unbounded class were considered equally (and highly) intentional (Bounded events, M = 5.81; Unbounded events, M = 5.71; $t\,(18) = 0.56$, $p > .250$). Thus any differences in the categorization of bounded or unbounded events could not be due to differences in intentionality.

We adopted three pictures of novel, simple objects and three pictures of novel substances from Li et al. (2009; see Figure 2). We chose these stimuli because in Li et al. (2009), speakers of English (but also other languages such as Japanese and Mandarin) rated the three pictures in the upper row of Figure 2 as "good objects" (with an average rating score above 6 on a 1-7 scale, where 1 was a good substance and 7 a good object) and the three pictures in the lower row as "good substances" (with an average rating score below 3 on the same scale).[1]

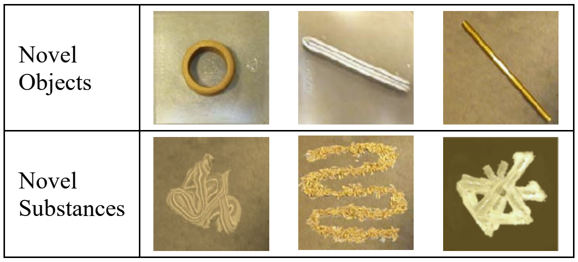

Figure 2: Picture stimuli in Experiment 1.

Procedure Participants were tested in small groups during an in-lab testing session. They were given a category identification task (adapted from Ji & Papafragou, 2020). In the training phase, participants were asked to watch a few videos and attend to those appearing within a *red* frame. Their task was to figure out what kind of videos could get a red frame. Participants were presented with 4 pairs of bounded and unbounded events. Each time, a single video was played in the center of the screen and was followed immediately by the other video within the pair (the order of bounded-unbounded events within pairs was counterbalanced). Participants were randomly assigned to one of two conditions. In the Bounded condition, the videos of bounded events in the training phase were surrounded by a red frame while their unbounded counterparts were surrounded by a black frame. In the Unbounded condition, the assignment of frame colors was reversed.

The testing phase consisted of two sessions. In the first, Video Session, participants saw 6 videos showing new events and for each one they were asked: "Could the video have a red frame or not?" Videos were displayed individually at the center of the screen. We arranged 6 pairs of bounded-unbounded events into 2 lists (see the testing block of Table 1). Each list contained one event from each pair, with equal numbers of bounded and unbounded events (n = 3). Whether the event was bounded or unbounded were counterbalanced across the two lists. Participants were randomly assigned to one of the two lists. In the second, Picture Session, participants were told: "What you found in the videos also works for pictures. Now you'll see a few pictures and decide whether a picture can have a red frame or not." They next saw 6 pictures of novel entities and, for each picture, they were asked: "Could the picture have a red frame or not?" Pictures were displayed individually at the center of the screen. We arranged the 6 pictures into 2 lists, each containing 3 object and 3 substance examples (see Figure 2). In both lists, the novel objects and substances were intermixed. One list began with a picture of a novel object, the other with a picture of a

---

[1] The last picture of a novel substance was an edited version. In the original version (Li et al., 2009), the sand-like substance had a very regular bow-tie shape. We split the "bow-tie" shape in halves and rearranged the two halves using Adobe Photoshop.

novel substance. Participants were randomly assigned to one of the two lists.

Results

We first examined responses in the Video Session. We coded a response as correct if participants could assign the red frame to a video of the target category, or reject the red frame for a video of the non-target category. For instance, participants in the Bounded condition were expected to give a *Yes* response to a test video of a bounded event, and a *No* response to a video of an unbounded event. The binary accuracy data were analyzed using multi-level mixed modeling with crossed intercepts for Subjects and Items (Baayen, et al., 2008; Barr, 2008). We coded Condition (Bounded vs. Unbounded) using centered contrast (-0.5, 0.5) and included it as the fixed factor. As shown in Figure 3, participants were better at identifying the Bounded (M = 78.5%) than the Unbounded (M = 66.0%) category (β = -0.67, SE = 0.33, z = -2.03, p = .042).

We then examined the responses in the Picture Session. Similar to the Video Session, we coded a response as correct if participants could extend the boundedness feature to static images and assign the red frame to a picture of the target category, or reject the red frame for a picture of the non-target category. For instance, participants in the Bounded condition were expected to give a *Yes* response to a picture of a novel object, and a *No* response to a picture of a novel substance. The binary data were submitted to a logit model examining the fixed effects of Condition (Bounded vs. Unbounded). As shown in Figure 3, unlike the Video Session, no significant difference was found between the Bounded (M = 66.7%) and the Unbounded (M = 74.3%) condition (β = -0.45, SE = 0.49, z = -0.91, p > .250).

In general, the average proportion of correct responses in both sessions in both conditions was significantly above chance level (ps < .02), suggesting that participants were able to identify the target event category in the first session and extend the (un)boundedness feature to the object domain in the following session.

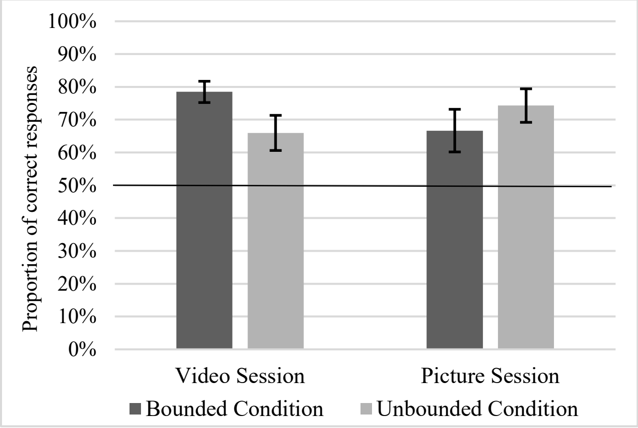

Figure 3: Proportion of correct responses in the two sessions of the testing phase in Experiment 1. Error bars represent ±SEM.

We further considered the relationship between the two sessions. We analyzed the accuracy data from both sessions using mixed-effects modeling to examine the fixed effects of Condition (Bounded vs. Unbounded), Session (Video vs. Picture) and their interaction. Only the Condition by Session interaction was found to be significant (β = 1.13, SE = 0.40, z = 2.86, p = .004): in the Bounded condition, performance got worse from the Video Session (M = 78.5%) to the Picture Session (M = 66.7%) (β = -0.67, SE = 0.28, z = -2.37, p = .018), but in the Unbounded condition, performance was not different between the two sessions (M = 66% in Video Session, M = 74.3% in Picture Session, β = 0.46, SE = 0.28, z = 1.66, p = .096).

Discussion

Experiment 1 first replicated two major findings from Ji and Papafragou (2020): participants could form categories of bounded vs. unbounded events when exposed to multiple different exemplars of each category. Furthermore, they were better at forming the category of bounded events compared to that of unbounded events, presumably because these events were "atoms" and, as such, could be individuated, compared and categorized more easily. This asymmetry between the bounded and unbounded event category is reminiscent of previous findings from cognitive tasks in the object domain that target types of individuals. For instance, 8-month-old infants can detect when two rigid, cohesive objects made of sand are replaced with one, but they fail to detect a change when two poured piles of sand are replaced with one pile (Huntley-Fenner, et al., 2002). Across development, entities that are considered individuated or 'atomic' behave differently from entities that are unspecified for these features: they are more easily quantified over and tracked (e.g., Pylyshyn, 2001; Pylyshyn & Storm, 1988; vanMarle & Scholl, 2003).

Going beyond previous findings, Experiment 1 also found that (un)boundedness in the event domain had a counterpart in the object domain: when viewers were invited to generalize event classes to non-event stimuli, they extended event boundedness to the property of being an object and event unboundedness to the property of being a substance. This finding supports the hypothesis that the contrast between bounded and unbounded events has similarities to the contrast between objects and substances, such that both contrasts are guided by a single notion of atomicity (see Bach, 1986; Jackendoff, 1991, among others).

Our last finding is worth highlighting: there was a cost of switching from categorizing events into boundedness classes (Video Session) to categorizing objects along similar lines (Picture Session) – but only in the Bounded condition. This might seem surprising: the connection of boundedness to objecthood might be expected to be more trackable and stable (since both involve atomicity) compared to the connection of unboundedness to substancehood (since both are unspecified for atomicity). We will account for this finding along with the pattern found in Experiment 2 in General Discussion.

Experiment 2

Experiments 1 invited participants to extend their reasoning about events to reason about objects, and found that participants did so in accordance to the (non)atomic properties of the entities involved. A further question concerned how easily people could access the parallel between the event and the object domain. Could viewers notice the bounded-unbounded contrast and generalize from one domain to another without any prior training or explicit instructions? In Experiment 2, we directly asked participants to match bounded vs. unbounded events with pictures of objects vs. substances. We predicted that, even when viewers freely inspect a single event, we should see evidence of their principled, atomicity-driven ability to link events and objects.

Method

Participants Forty-eight adults (30 female, 18 male; M_{age} = 19.2, age range: 18.0 – 22.5) participated in the experiment. All were undergraduates at a major research university on the East Coast of the United States and received course credit for participation.

Stimuli The video and picture stimuli were identical to Experiment 1. The 10 pairs of videos showing bounded and unbounded events were separated into 2 lists. Each list contained one video from each pair. We counterbalanced whether the event was bounded or unbounded across lists, resulting in 5 bounded and 5 unbounded events per list. Each list was further arranged into two presentation orders. In one order, the 5 bounded events appeared before the 5 unbounded events; in the other order, unbounded events appeared before bounded ones.

The 6 pictures (see Figure 2) formed 9 object-substance pairs. Throughout the experiment, all of the pairs were used and one pair was used twice. The picture pairs were arranged in a pseudorandomized order such that the same picture would not appear in a row for more than twice.

We displayed the picture pairs and the videos as Figure 4 shows. In each trial, a video appeared on the left of the screen and a pair of pictures appeared on the right. Whether a novel object or a novel substance appeared on the top was counterbalanced across the trials.

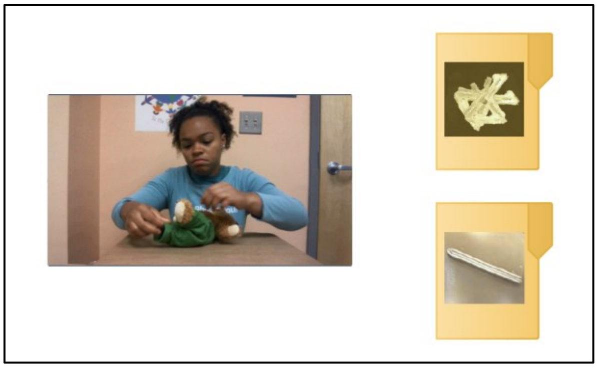

Figure 4: Display of the stimuli in Experiment 2.

Procedure At the beginning of the experiment, participants were told that they were going to organize some videos and put them into different folders. Each time a video was played and there were two folders available. The task was to decide which folder the video should go into.

Results

A response was coded as correct if a picture of a novel object was chosen for a bounded event, or a picture of a novel substance was chosen for an unbounded event. The binary accuracy data were submitted to a logit mixed-effects model examining the fixed effects of Event Boundedness (Bounded vs. Unbounded). As shown in Figure 5, participants were unexpectedly more accurate in choosing a picture of a substance for an unbounded event (M = 72.5%), compared to choosing a picture of an object for a bounded event (M = 62.9%) (β = 0.44, *SE* = 0.20, z = 2.24, p = .025). [2] Performance was significantly different from chance level across both types of events (*ps* < .025).

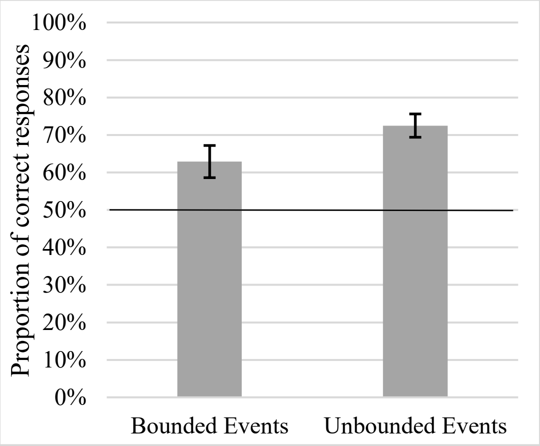

Figure 5: Proportion of correct responses in Experiment 2. Error bars represent ±SEM.

---

[2] No effect of order was detected (p > .250): the proportion of correct responses was 67.5% when bounded event trials appeared first and 69.5% when unbounded event trials appeared first. There was no significant interaction between event type and order as well (p = .204).

Discussion

Results from Experiment 2 showed that viewers could draw a link between bounded events and objects, and between unbounded events and substances from the beginning. Different from Experiment 1 where participants were exposed to 4 minimal pairs of bounded-unbounded events in a training phase, in this experiment people had to derive the contrast in atomicity from just two pictures. Therefore, the results demonstrated a strong and easily accessible parallel between the event and object domain, and offered further evidence for the conclusion that the representational units recruited by event and object cognition are similar. Similar to Experiment 1, an unexpected pattern was detected: it was easier to connect unboundedness to substancehood than boundedness to objecthood. We return to the significance of this last finding below.

General Discussion

In the present study, we took as our point of departure the well-established idea that boundedness is an organizing property of linguistic event representations and instantiates a notion of atomicity that also characterizes the linguistic semantics of objects (Bach, 1986; Jackendoff, 1991; see also Truswell, 2019; Wagner & Carey, 2003; Wellwood et al., 2018). On this proposal, in language, bounded (or telic) predicates resemble count nouns because they both are atomic with well-defined boundaries. By contrast, unbounded (or atelic) phrases resemble mass nouns in that they are unspecified for atomic features. We sought to test the idea that there should be a strong homology between the cognitive mechanisms underlying bounded/unbounded event construals on the one hand and object/substance entity construals on the other. If true, this idea could link the unit of both event and object representation to a single, foundational notion of atomicity.

Our experiments offered direct evidence in support of this idea. In Experiments 1, we found that, after forming categories of bounded or unbounded events, viewers could successfully extend those categories to instances of objects or substances respectively. In Experiment 2, viewers were able to spontaneously draw connections between events and objects in the absence of prior training or explicit instructions. Thus the cognitive representations of bounded/unbounded events and objects/substances align in ways that could be plausibly underscored by a common atomicity property.

A somewhat unexpected aspect of our results was that the advantage of the bounded event category over the unbounded one did not transfer to categorization of entities (Experiment 1). Meanwhile, the connection between bounded events and objects was not stronger, (and tended to be weaker) than that between unbounded events and substances (Experiment 2). We do not believe that these results showed that the non-atomic property was more discoverable or stable than the atomic property. There could be several reasons that result in these seemingly contradictory patterns. Participants could have come up with more specific conjectures about bounded events (e.g., regular motions) while the novel objects may not have all these features. Furthermore, the novel objects seemed to have their own function, which was unfamiliar, and unrelated to the videos. By contrast, participants' hypotheses about the class of unbounded events might be easier to connect to a new class of stimuli: as long as a picture lacked structure, neatness, etc., for instance, people would be able to relate it to an unbounded event.

The present data leave open several directions for future work. One important question is to ascertain how conceptual representations of boundedness arise in the mind. One possibility is that – in accordance with our hypothesis – atomicity could be a cognitive primitive and conceptual boundedness precedes and structures the linguistic encoding of boundedness. Alternatively, the direction of causation might be the reverse: the conceptual signature of boundedness might arise because of familiarity with the way boundedness is encoded in the viewer's language. Only the first hypothesis predicts that non-linguistic event boundedness would be conceptualized in similar ways cross-linguistically. We plan to test this prediction in future cross-linguistic work.

Acknowledgments

This material is based upon work supported by the Beijing Institute of Technology Research Fund Program for Young Scholars (Y.J.) and National Science Foundation Grant 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,* 390–412.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy, 9,* 5–16.

Baillargeon, R. & Wang, S. (2002). Event categorization in infancy. *Trends in Cognitive Science, 6*(2), 85-93.

Barner, D., Wagner, L., & Snedeker, J. (2008). Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns. *Cognition, 106*, 805–832.

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

Buchsbaum, D., Griffiths, T.L., Gopnik, A., & Baldwin, D. (2015). Learning from actions and their consequences: Inferring causal variables from continuous sequences of human action. *Cognitive Psychology, 76*, 30–77.

Cooper, R. P. (2021). Action production and event perception as routine sequential behaviors. *Topics in Cognitive Science, 13,* 63–78.

Elman, J., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*, 252–291.

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

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of distinct entities. *Cognition, 85*, 203–221.

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

Jackendoff, R. (1991). Parts and boundaries. *Cognition, 41*, 9–45.

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

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

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

Levine, D., Hirsh-Pasek, K., Pace, A., & Golinkoff, R. (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.

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: the nature of language effects on entity construal. *Cognitive Psychology, 58*(4), 487–524.

Malaia, E. Renaweera, R., Wilbur, R., & Talavage, T. (2012). Event segmentation in a visual language: Neural bases of processing American Sign Language predicates. *Neuroimage, 19,* 4094–4101.

Mourelatos, A. P. (1978). Events, processes, and states. *Linguistics and Philosophy, 2*(3), 415–434.

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology, 61*(2), 63–86.

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

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.

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

Pulverman, R., Song, L., Hirsh-Pasek, K., Pruden, S. M., & Golinkoff, R. M. (2013). Preverbal infants' attention to manner and path: Foundations for learning relational terms. *Child Development, 84*(1), 241–252.

Pylyshyn, Z. W. (2001). Visual indexes, perconceptual objects, and situated vision. *Cognition, 80*, 127–158.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197.

Radvansky, G., & Zacks, J. (2014). *Event cognition*. Oxford University Press.

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

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science, 13*(6), 493–498.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distorted memory in a matter of seconds. *Cognition, 121,* 409–415.

Strickland, B., & Scholl, B. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General, 144*(3), 570–580.

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences of the United States of America, 113*(9), 5968–5973.

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

Taylor, B. (1977). Tense and continuity. *Linguistics and Philosophy, 1,* 199–220.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*(1), 121–125.

Truswell, R. (Ed.). (2019). *The Oxford handbook of event structure*. Oxford University Press.

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

Van Marle, K., & Scholl, B. J. (2003). Attentive tracking of objects vs. substances. *Psychological Science*, *14*, 498–504.

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

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition, 90,* 163–191.

Wellwood, A., Hespos, S. J., & Rips, L. (2018). The object : substance :: event : process analogy. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy*, Volume 2. Oxford University Press.

Wittenberg, E., & Levy, R. (2017). If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language, 94,* 254–271.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111.

Woodward, A. L., & Cannon, E. (2013). Online action analysis: Infants' anticipation of others' intentional actions. In M. Rutherford & V. Kuhlmeier (Eds.) *Social perception*. MIT Press.

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