Events and objects are similar cognitive entities

Anna Papafragou a, Yue Ji b, *

a Department of Linguistics, University of Pennsylvania, 3401-C Walnut St., Philadelphia, PA 19104, USA
b Department of English, School of Foreign Languages, Beijing Institute of Technology, No. 5 South Street, Zhongguancun, Haidian District, Beijing 100081, China

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” that have well-defined boundaries, contain discrete minimal parts and cannot be divided arbitrarily. 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 demonstrate for the first time the parallels in the perceptual-cognitive representation of events and objects even in entirely non-linguistic tasks. Specifically, after viewers form categories of bounded or unbounded events, they can extend the category to objects or substances respectively (Experiments 1 and 2). Furthermore, in a training study, people successfully learn event-to-object mappings that respect atomicity (i.e., grouping bounded events with objects and unbounded events with substances) but fail to acquire the opposite, atomicity-violating mappings (Experiment 3). Finally, viewers can spontaneously draw connections between events and objects without any prior training (Experiment 4). 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.

1. Introduction

Humans are surprisingly adept at interpreting what is happening around them, and rapidly segmenting and organizing dynamic experience into coherent event construals. Such event construals offer a starting point for assembling a linguistic description of the event during speaking (Konopka & Brown-Schmidt, 2014; Levelt, 1989; cf. Clark, 1973; Jackendoff, 1996). Similarly, core aspects of event construals constrain the child’s hypotheses about what words mean during language learning (Gleitman, 1990; Pinker, 1989). However, the precise format of event construals and their mapping to language have remained elusive, partly because research on how people mentally segment and perceive events (see Radvansky & Zacks, 2014; Shipley & Zacks, 2008, for reviews) has largely proceeded separately from analyses of how events are encoded in language (see Binnick, 2012; Truswell, 2019, for reviews).

Here we aim to throw light onto the unit of event cognition and eventually contribute to our understanding of how event units are transformed into language. Bringing together two dominant but separate lines of event research, one on perceptual-cognitive mechanisms of event representation, and the other on logico-semantic event structure, we propose that the cognitive unit of event representation is similar to the unit of object representation. In the remainder of the Introduction, we outline relevant theoretical and empirical background before turning to our own studies.

* Corresponding author at: Department of English, School of Foreign Languages, Beijing Institute of Technology, Beijing 100081, China.
E-mail addresses: anna4@sas.upenn.edu (A. Papafragou), jiyue@bit.edu.cn (Y. Ji).

https://doi.org/10.1016/j.cogpsych.2023.101573
Received 14 July 2022; Received in revised form 25 April 2023; Accepted 2 May 2023
Available online 11 May 2023
0010-0285/© 2023 Elsevier Inc. All rights reserved.
1.1. Event structure: cognitive perspectives

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 (Radavansky & Zacks, 2014). Rich evidence from event segmentation studies shows that event boundaries (especially event endpoints) are important for event perception and comprehension (e.g., boundaries are recalled and described more accurately than non-boundaries; Huff et al., 2012; Newson & Enquist, 1976; Swallow et al., 2009; cf. also Lakusta & Landau, 2005, 2012; Levine et al., 2017; Papafragou, 2010; Regier & Zheng, 2007). Other frameworks of event representation have addressed broad parameters of complex event knowledge as opposed to the internal structure of events themselves (e.g., Cooper, 2021; Elman & McRae, 2019; Zwaan & Radvansky, 1998). Experimental work has offered rich insights on how infants, children and adults conceptually represent dynamic events but has tended to target 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; Wolff, 2007; and intentionality – Mathis & Papafragou, 2022; Tomasello & Carpenter, 2007; Woodward & Cannon, 2013; among many others).

A topic that has received much less discussion in the above-mentioned works on event segmentation and event structure is how people encode the representational unit within event boundaries (i.e., “what happens” between the time an event begins and ends). In general, the representational profile of an event as a class has only been the topic of a handful of studies (see, e.g., Altmann & Ekves, 2019, for an explicit proposal that defines events as intersecting changes of state in an affected entity; cf. also Lee & Kaiser, 2021; Sakarias & Flecken, 2019). An important goal of general models of event representation is thus to specify how the process of mentally assembling event units and their boundaries yields an understanding of the organization of events, including the nature of event boundariness.

1.2. Event structure: linguistic perspectives

According to a separate linguistic and philosophical tradition on events that goes back to Aristotle (see Filip, 2012), telicity in language describes a dynamic situation as either a bounded or an unbounded event (Binnick, 2012; Jackendoff, 1991; Mourelatos, 1978; Parsons, 1990; Truswell, 2019; Van Hout, 2016; Vendler, 1957). 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 distinct endpoint – the moment when Sam is in the house. By contrast, the atelic sentence Sam approached the house encodes an experience as an unbounded event: this event lacks distinct stages since “any part of the process is of the same nature as the whole” (Vendler, 1957, p. 146) – each moment of the action can still be described as an event of approaching the house. Unbounded events have no specified endpoint and may terminate at an arbitrary moment. Several studies have confirmed that telicity conjures different cognitive perspectives on the temporal structure of events during language comprehension (Barner et al., 2008; Malaia et al., 2012; Strickland et al., 2015; Wagner & Carey, 2003; Wellwood et al., 2018a).

Telicity in language is not simply a cornerstone of the logical and linguistic analysis of events but has been widely assumed to be grounded in non-linguistic event representations (Filip, 1993). A recent series of studies connect telicity to boundedness, a cognitively active feature of (non-linguistic) event structure, despite the absence of the latter from current cognitive event frameworks (Ji & Papafragou, 2020a, 2020b, 2022). In one of these studies (Ji & Papafragou, 2020a), viewers were exposed to pairs of videos showing a bounded and an unbounded event, such as dress a teddy bear (bounded) vs. pet a teddy bear (unbounded), or eat a pretzel (bounded) vs. eat cheerios (unbounded); one member of the pair was marked. Participants were able to extract the marked category of events (bounded vs. unbounded) from exposure to these paired exemplars and generalize it to new event items. Further experiments showed that viewers distinguished between boundedness (the availability of an inherent endpoint) and event completion (the actual realization of the inherent endpoint, Ji & Papafragou, 2020; cf. Ji & Papafragou, 2019, for a replication with 4–5-year-olds). Moreover, viewers identified bounded and unbounded classes even when linguistic encoding was suppressed, removing the possibility that the categorization results were due to online encoding of the events in terms of (a)telic English sentences (Ji & Papafragou, 2020). Together, these findings show that boundedness is an organizing property of non-linguistic event representations even when people do not have to produce or comprehend event descriptions.

1.3. Beyond events: counterparts to boundedness in the object domain

Within linguistics, a number of commentators have argued that the notion of telicity in the semantics of verbal predicates has a counterpart in the mass/count distinction in the semantics of nominals, and that this correspondence is supported by the notion of atomicity (Bach, 1986; Jackendoff, 1991; Taylor, 1977; cf. Champollion, 2015, 2017; Filip, 2012, 2019; Wellwood et al., 2018a). 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; cf. 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.

Studies of how linguistic form connects to conceptualization support the parallel between the cognitive representations of events and objects (e.g., Barner et al., 2008; Kuhn et al., 2021; Prasada et al., 2002; Wellwood et al., 2018a, 2018b; see also Maguire et al., 2011; Sharon & Wynn, 1998; Wood & Spelke, 2005). For instance, adults preferred to describe a novel, regularly-shaped piece of material using count syntax (There is a blicket), but an irregularly-shaped piece (i.e., a non-canonical individual) using mass syntax (There is blick; Prasada et al., 2002). These differences in conceptualization affect the comprehension of sentences with more: 3-year olds judged sentences with more blick (mass syntax) to be true based on a number comparison only when blick clearly applied to solid objects but uniformly judged sentences with more blicks (count syntax) to be true on the basis of number, regardless of what counted as blick (Barner & Snedeker, 2005). In the event domain, adults shifted the relevant dimension for evaluating sentences with do more V-ing (mass syntax) depending on the boundedness profile of the verb (when the verb was telic, as in do more kicking, the dimension was number of kicks, but when the verb was atelic, as in do more dancing, the dimension of evaluation was temporal duration of the dancing); in contrast, do more Vs (count syntax) uniformly led participants to evaluate the sentence on the basis of number of actions (Barner et al., 2008). In another study (Wellwood et al., 2018a), observers tended to describe images with “natural” spatial breaks and motions with “natural” temporal breaks using plural count nouns and telic verb phrases respectively but choose mass nouns and atelic verb phrases for unnaturally-divided, thus non-atomic images and motions. More recently, Kuhn et al. (2021) showed that non-signers tended to associate signs that have a gestural boundary or acoustic stops (e.g., /p/, /v/, /k/) that have an auditory boundary with telic verbs and count nouns. In the next section, we test the connection between event and object atomicity using exclusively non-linguistic probes of entity construals.

1.4. The present study

Our goal in the present paper is to probe whether temporally extended entities (bounded/unbounded events) and spatially extended entities (objects/substances) align in non-linguistic cognition in accordance with abstract considerations of atomicity. Two points of clarification are in order about the choice of “bounded vs. unbounded events” to refer to members of the event domain, and “objects vs. substances” to refer to members of the object domain. First, both sets of terms refer to human conceptualization rather than objective classes of entities in the world. A single temporal entity in the world can often be described as either a bounded or unbounded event (e.g., play the Moonlight Sonata vs. play the piano), just like a single spatial entity can be described as either an object or a substance (e.g., a tree vs. wood). Despite the flexibility of descriptions and construals, several factors might bias a viewer in favor of one or the other option. The present study focuses on instances that are strongly biased in the direction of one construal rather than the other (but see Yurgun et al., 2022; Wagner & Carey, 2003). Second, the present distinctions give only one possible example of temporal and spatial entities that differ in their individuation profile (and correspond imperfectly with the telic/atelic and count/mass distinctions in language; e.g., Barner & Snedeker, 2005). We use them here given the emphasis on these distinctions in the cognitive and developmental literature on individuation.

Our study brings widely held architectural assumptions about the nature of spatial (object) and temporal (event) entities from linguistics and philosophy to bear on experimentally testable cognitive theories of how these entities are represented in cognition. Building on our hypothesis that boundedness structures event units in cognition, we posit that bounded events resemble objects (they are “atoms” that can be individuated and compared to each other) and unbounded events resemble substances (they are unspecified for atomic features). We then test this proposal in a series of four interlocking experiments.

Evidence that the cognitive representations of bounded/unbounded events and objects/substances align in non-linguistic cognition is theoretically significant for multiple reasons. First, by throwing light onto the mental units of eventhood and objecthood, such evidence has the potential to enrich theories of event cognition. Currently, on many such cognitive theories (e.g., classic segmentation theories as in Zacks et al., 2007), a correspondence between event and object structure is entirely unexpected. Second, such evidence can bear on the scope and explanatory power of semantic theories of lexical aspect and telicity: if similarities between mental properties of events and objects belong (at least in part) to cognitive architecture, then they would fall outside the grammar of natural languages and the scope of linguistic theories should be correspondingly adjusted (Filip, 2012). Third, evidence of a link between bounded/unbounded events and objects/substances would illustrate natural connections between event cognition and language that should themselves have implications for language use and acquisition: being able to categorize an event as having an inherent boundary or a spatial entity as being individuated can offer a conceptual starting point for planning a sentence about the event or object (Levitt, 1989), as well as scaffold the way learners acquire the tools for encoding events and objects in their native tongue (cf. Gleitman, 1990; Pinker, 1989).

Our empirical investigations proceed as follows. In Experiments 1 and 2, we build on the boundedness studies by Ji and Papafragou (2020a) to test whether people can extend the category of event (un)boundedness to the corresponding distinction between objects and substances. In Experiment 3, we ask whether people can match events to objects after being trained on either natural mappings (e.g., bounded events to objects, unbounded events to substances) or unnatural mappings (e.g., bounded events to substances, unbounded events to objects). If the present line of reasoning is correct, only training with natural cross-domain mappings should result in successful performance. Finally, in Experiment 4, we ask whether viewers can draw connections between events and objects in the absence of any prior training.
2. Experiment 1

In this experiment, we asked whether viewers would be able to extend the (un)boundedness feature from the event domain to a corresponding distinction in the object domain. In a category identification task (adapted from Ji and Papafragou, 2020a), 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.

2.1. Method

2.1.1. Participants

Forty-six adults (28 female, 18 male; M_age = 19.2, age range: 18.0 – 21.5) participated in the experiment. All were undergraduates recruited from the University of Delaware and received course credit for participation. For this and all subsequent studies in this paper, informed consent was obtained prior to participation. The sample size was decided based on the calculated power of Experiment 1 in Ji and Papafragou (2020a) that had the same design with 40 participants. The power analysis was based on the reported mixed-effects model (Ji and Papafragou, 2020a, p. 5) with Condition (Bounded vs. Unbounded) as the fixed effect of interest, using the simr package in R (Green & MacLeod, 2016). The estimated effect size for Condition was -1.4, and the power for this predictor was 0.92, suggesting that a sample size of 40 participants would be adequate to achieve 80% power at α = 0.05. The sample sizes of the following experiments were decided in the same way.

2.1.2. Stimuli

2.1.2.1. Video stimuli. We used the same 20 pairs of videos as in Ji and Papafragou (2020a). All videos involved the same girl doing a familiar everyday action in a lab room. The videos began with the girl picking up an object or a tool from the desk and ended with her hands away from the object or tool. Paired videos had the same duration (range: 4.4 s – 12.0 s, M = 7.8 s, SD = 2.4), and showed a bounded and an unbounded event (see Table 1 and Fig. 1). The bounded-unbounded contrast was due to either the nature of the action or the nature of the affected object (e.g., Filip, 2004; Tenny, 1987). For half of the videos, paired bounded and unbounded events had the same object but differed in the action performed on the object: the bounded event involved an action that caused a clear and temporally demarcated change of state in the object (e.g., stack a deck of cards) while its unbounded counterpart did not have such a change (e.g., shuffle a deck of cards). For the other half of the videos, paired bounded and unbounded events had the same action but differed in the nature of the affected object: the bounded event involved a single object (e.g., tie a knot) but its unbounded counterpart involved either an unspecified plurality or objects or a mass quantity (e.g., tie knots). We tried to match the degree of visual repetition within paired videos. For instance, in the (bounded) event of eating a pretzel, the actor took two bites in order to finish the pretzel; in the (unbounded) event of eating cheerios, the actor also ate twice, each time placing her hand into the bag and taking out a quantity of cheerios. (Indeed, when later describing the meaning exemplified in the bounded vs. unbounded class, participants did not mention repetition for either class; see section 2.2.3.)

Two types of norming studies were used to confirm the placement of the stimuli within the bounded vs. the unbounded class. First, a separate group of 40 participants provided judgments about the temporal structure of the stimuli. The 20 pairs of events in Table 1 were split into two lists, such that each list included only one video from each pair, and had an equal number (N = 10) of bounded and unbounded events. Participants were randomly assigned to one of the two lists. After watching each video clip, participants answered a Yes/No question: “Does it make sense to think of the action in the video as something with a beginning, midpoint and specific endpoint?” Videos of bounded events elicited Yes responses 87% of the time (i.e., 348 out of 400 responses) while videos of unbounded events did so only 21.5% of the time (i.e., 86 out of 400 responses) (a significant difference, \( t(39) = 20.33, p < .0001 \)). Second, a new group of 18 native speakers of English described the video clips (Ji & Papafragou, 2020a). Their descriptions underwent linguistic tests for boundedness (e.g., telic descriptions can be modified by delimited temporal phrases such as in an hour while atelic descriptions go along with durative temporal phrases such as for an hour; see Dowty, 1979; Vendler, 1957). Videos of bounded events elicited telic descriptions 98.2% of the time, including change-of-state predicates (e.g., stack/pile up a deck of cards) or quantified noun phrases (e.g., tie a knot); stimuli of unbounded events elicited atelic verb phrases with verbs of activity (e.g., shuffle a deck of cards) or unquantified noun phrases (e.g., tie knots) 92.8% of the time (Ji & Papafragou, 2020a). There was no significant difference between the two event types in terms of whether they elicited the corresponding aspectual distinctions in the production task (t (17) = 1.84, p = .083).

Two additional rating studies were conducted to evaluate possible features of the stimuli that might affect event categorization. First, we assessed the degree of intentionalty of all the videos used in the experiment. A new group of 20 participants rated how intentional the action in each video looked on a scale from 1 (totally unintentional) to 7 (intentional) (Ji & Papafragou, 2020a). The degree of intentionality did not differ between bounded (M = 5.67) and unbounded (M = 5.62) stimuli (t (19) = 1.34, p = .195), indicating that instances within the bounded and unbounded class were considered equally (and highly) intentional. Second, we

---

1 Notice that these responses were not at ceiling for either class, probably because any event is multiply interpretable (and, on a literal interpretation of the question, within the videos themselves the action started and stopped clearly, i.e., it had a beginning and an end). Overall, however, there was an overwhelming preference within each item to be classified as we expected, and whatever variability of answers existed on either side of the boundedness divide was stable across items.
assessed the degree of visual similarity of the event instances used in the training phase in the bounded vs. unbounded class. All possible pairwise combinations of the 8 videos of bounded events in the training phase (see Table 1) were intermixed with all possible pairwise combinations of the 8 videos of unbounded events. A different group of 20 people rated the degree of visual similarity for each pairwise combination on a scale from 1 (least similar) to 7 (most similar). The average rating for pairs of bounded events ($M = 2.38$) did not differ from those of unbounded events ($M = 2.52$) ($t(19) = -1.43, p = .169$). Thus any differences in the categorization of bounded or unbounded events could not be due to differences in intentionality or visual similarity of exemplars in the bounded vs. unbounded class.

### 2.1.2.2. Picture stimuli

We adopted three pictures of novel, simple objects and three pictures of novel substances from Li et al. (2009; see Fig. 2). We chose these stimuli for two reasons. First, 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 Fig. 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

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

Note: To avoid the possibility that people formed categories simply based on the specific event pairs used in the training phase, we rotated videos used across the training and the testing phase.

Fig. 1. Pairs of video stimuli in Experiment 1: (a) put up one’s hair (bounded) vs. (b) scratch one’s hair (unbounded).
average rating score below 3 on the same scale). Second, since our event stimuli already involved familiar objects or substances, using pictures of novel things could reduce the possibility that observers would make incidental real-world connections between the video and the picture stimuli.

2.1.3. Procedure

Participants were tested in small groups during an in-lab testing session. They were given a category identification task (adapted from Ji & Papafragou, 2020a).

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 8 pairs of bounded and unbounded events, among which 4 pairs showing the bounded-unbounded contrast in the Action were intermixed with 4 pairs showing the contrast in the Affected object (see the training block of Table 1). 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 8 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 8 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 = 4). Whether the event was bounded or unbounded and whether the source of boundedness was the action or the affected object 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 Fig. 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 novel substance. Participants were randomly assigned to one of the two lists.

In a last, surprise testing phase, participants watched 4 videos of new events and were unexpectedly asked about the black frame (“Could the video have a black frame or not?”). This question was meant to probe whether participants had formed any hypotheses about the secondary category present, even though it was not the target. We arranged 4 pairs of videos into 2 lists (see the last block of Table 1) and used the same counterbalancing as in the Video Session of the testing phase.

At the end of the experiment, participants were asked to write down their answers to two open questions: (1) What kind of videos can have a red frame? (2) What kind of pictures can have a red frame? The two questions were used as an additional source of information about the category that participants had just formed.

---

2 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.
2.2. Results

2.2.1. Performance in the testing phase

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 included Condition (Bounded vs. Unbounded) as the fixed factor (dummy coded as 0, 1)). As shown in Fig. 3 and Table 2, there was an effect of Condition (β = -1.29, z = -3.41, p < .001), i.e., participants were better at identifying the bounded event category (M = 90.2%, SE = 0.017) compared to the unbounded category (M = 73.4%, SE = 0.047). Performance in both conditions was significantly different from chance level (ps < 0.001), suggesting that participants were able to identify both the bounded and the unbounded category.

To further bolster this conclusion, we ran two linear regressions to investigate whether the Yes/No answers to the question about event temporal structure and the ratings of the degree of intentionality obtained from other groups in prior norming could significantly predict participants’ categorization of the event items. The analyses revealed that the proportion of Yes responses to the norming question “Can the action in the video be considered as something with a beginning, midpoint and specific endpoint?” explained a significant portion of variance in the proportion of testing responses that categorized an event as bounded ($R^2 = 0.864, F(1, 30) = 190.1, p < .0001$). In other words, as we expected, events that were more likely to be previously judged as having specific temporal boundaries tended to be identified as bounded. By contrast, the ratings of the degree of intentionality during norming did not predict whether an event would be identified as bounded ($R^2 = 0.096, F(1, 30) = 3.2, p = .084$).

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 effect of Condition (Bounded vs. Unbounded). As shown in Fig. 3 and Table 3, performance did not differ between the Bounded (M = 79.7%, SE = 0.036) and Unbounded condition (M = 81.9%, SE = 0.044) ($p > .250$). Performance in both conditions was significantly different from chance level (ps < 0.001), suggesting that participants could extend their reasoning from videos to pictures.

We further considered the relationship between the two sessions. Overall, the accuracy of the Video Session was positively correlated to that of the Picture Session ($r = 0.549, p < .001$). 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. As shown in Table 4, a fixed effect of Condition was detected (β = -1.25, z = -3.08, p = .002), showing that participants overall were better at identifying the Bounded category (M = 85.0%, SE = 0.073) compared to the Unbounded one (M = 77.6%, SE = 0.088). There was also an effect of Session (β = 0.93, z = -2.55, p = .011), i.e., performance became worse from the Video (M = 81.8%, SE = 0.028) to the Picture Session (M = 80.8%, SE = 0.028). This suggests a cost of switching from one domain to another. Last, there was a significant interaction between Condition and Session (β = 1.50, z = 3.38, p < .001). In the Bounded condition, participants performed worse in the Picture Session (M = 79.7%, SE = 0.036) than in the Video Session (M = 90.2%, SE = 0.017) (β = -0.87, z = -2.39, p = .017). By contrast, in the Unbounded condition, there was no significant difference between the Picture (M = 81.9%, SE = 0.044) and the Video Session (M = 73.4%, SE = 0.047) (β = 0.60, z = 1.92, p = .055).

We also examined whether the Source of Boundedness (Action vs. Affected Object), and Stimuli Boundedness (i.e., whether a test trial showed a bounded or an unbounded event) would influence accuracy. We built up our logit model in a bottom-up fashion: the Source of Boundedness, Condition, and their interactions were added incrementally to the model to see whether the model fit was improved (assessed using $F(1, 30) = 6.52, p < .0001$). In the Bounded condition, there was a significant difference between the Picture (M = 90.2%, SE = 0.017) and the Video Session (M = 81.9%, SE = 0.044) ($p < .0001$). In the Bounded condition, participants performed worse in the Picture Session (M = 79.7%, SE = 0.036) than in the Video Session (M = 90.2%, SE = 0.017) (β = -0.87, z = -2.39, p = .017). By contrast, in the Unbounded condition, there was no significant difference between the Picture (M = 81.9%, SE = 0.044) and the Video Session (M = 73.4%, SE = 0.047) (β = 0.60, z = 1.92, p = .055).

Recall that the class of bounded events was at least in some cases explicitly set up to involve a demarcated object change compared to the unbounded class (section 2.1.2.1). Could this difference, if confirmed, drive the present boundedness categorization? To investigate this possibility, we asked a new group of participants (N = 14) to indicate whether each of the 20 videos showed a change in an object on a Likert scale from 1 (no change at all) to 7 (substantial change). Perhaps unsurprisingly, bounded events were found to involve higher object change (M = 4.7) compared to unbounded events (M = 3.5; t (13) = 6.52, p < .0001). However, in a linear regression, these ratings of object-state change were not a significant predictor of the tendency to identify an event as bounded in the current testing phase ($R^2 = 0.079, F(1, 30) = 2.6, p = .118$). We conclude that object-state changes are a natural but not the only way of delimiting an event, and do not capture broad boundedness computations.

We reanalyzed the performance in the Picture Session using only participants who had at least 6 correct responses out of the 8 trials in the Video Session (i.e., those who had already correctly identified the (un)boundedness target category). This analysis included all of the 23 participants in the Bounded condition but only 13 of the 23 participants in the Unbounded condition. Participants in the Unbounded condition (M = 92.3%) performed better than those in the Bounded condition (M = 79.7%) (β = 1.14, z = 2.26, p = .024) but in both cases performance was reliably higher than chance level.

By-subject and by-item random slopes for the effect of Session (Video vs. Picture) were not included in the final model because these models did not converge. The same treatment of random slopes can be found in similar work on the interface between event language and event cognition (e.g., Kuhn et al., 2021).
2.2.2. Performance in the surprise testing phase

The responses to surprise questions were analyzed in the same way as the responses in the main testing phase. No significant difference was found between the Bounded (M = 67.4%, SE = 0.048) and Unbounded condition (M = 64.1%, SE = 0.101) (p > .250; see Table 5). Participants performed at levels different from chance in both conditions (ps < .05), suggesting that they also formed a generalization about the non-target category.

Table 2
Fixed effect estimates for multi-level model of event category identification in the Video Session of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.82</td>
<td>0.26</td>
<td>6.90***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>-1.29</td>
<td>0.38</td>
<td>-3.41***</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition. * p < .05, ** p < .01, *** p < .001.

Table 3
Fixed effect estimates for multi-level model of category identification in the Picture Session of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.60</td>
<td>0.22</td>
<td>7.22***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>0.17</td>
<td>0.39</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition. * p < .05, ** p < .01, *** p < .001.

Table 4
Fixed effect estimates for multi-level model of category identification in the testing phase of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.48</td>
<td>0.34</td>
<td>7.38***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>-1.25</td>
<td>0.41</td>
<td>-3.08**</td>
</tr>
<tr>
<td>Session (Video vs. Picture)</td>
<td>-0.93</td>
<td>0.36</td>
<td>-2.55*</td>
</tr>
<tr>
<td>Condition*Session</td>
<td>1.50</td>
<td>0.44</td>
<td>3.38***</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition + Session + Condition: Session. * p < .05, ** p < .01, *** p < .001.

2.2.2. Performance in the surprise testing phase

The responses to surprise questions were analyzed in the same way as the responses in the main testing phase. No significant difference was found between the Bounded (M = 67.4%, SE = 0.048) and Unbounded condition (M = 64.1%, SE = 0.101) (p > .250; see Table 5). Participants performed at levels different from chance in both conditions (ps < .05), suggesting that they also formed a generalization about the non-target category.

Table 5
Fixed effect estimates for multi-level model of category identification in the surprise testing phase of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.80</td>
<td>0.28</td>
<td>2.89**</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>-0.15</td>
<td>0.38</td>
<td>-0.40</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition. * p < .05, ** p < .01, *** p < .001.
2.2.3. Answers to the open questions

We had 45 answers to each of the open questions (one participant did not answer them). For the question “What kind of videos can get a red frame?” 23 answers (51%) referred to the organization of the stimuli (e.g., the bounded event category was described as “organized” or “orderly” while the unbounded category was “disorganized” or “out of order”); 12 answers (27%) focused on the neatness of the actions (e.g., bounded events were “clean” and “neat” while unbounded ones were “messy” and “chaotic”); 3 answers (7%) made reference to goal or purpose (e.g., bounded events were “purposeful” while unbounded events did not “have an end goal”). In addition, 5 answers (10%) were a mixture of “organization”, “neatness” and “goal”. Only 2 responses (5%) failed to provide a summary but described specific event examples. These results suggest that participants could extract abstract features from the video stimuli to paraphrase the bounded or unbounded category as we expected.

The answers to the second question (“What kind of pictures can get a red frame?”) had some overlap with the answers to the first question: 15 answers (33%) made reference to the structure or organization of the stimuli (e.g., novel objects were “structured” while novel substances “had no pattern”, or “structure”); 10 answers (22%) referred to neatness (e.g., novel objects were “neat”, and “clean” while novel substances were “messy”, “jumbled”). Moreover, 20 answers (45%) commented more specifically on the shapes and lines (e.g., novel objects were “symmetrical” with “a given shape” while novel substances were “unsymmetrical” with “irregular shapes” or “rough edges”). These comments were in line with previous findings suggesting that observers understood entities based on their shapes and tended to consider entities with a non-arbitrary shape as objects (Li et al., 2009; Prasada et al., 2002).

2.3. Discussion

Experiment 1 replicated two major findings from Ji and Papafragou (2020a): participants could form categories of bounded vs.
unbounded events when exposed to multiple different exemplars of each category (even when these categories were not the main objective of the task; cf. surprise testing phase); 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. Importantly, the observed patterns were predicted by explicit ratings of whether the event had distinct stages such as a beginning, midpoint and endpoint but not by other ratings such as the degree of intentionality or change in the object.

Going beyond previous findings, Experiment 1 further 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 selectively 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 is similar to the contrast between objects and substances. This similarity was reflected in participants’ abstract paraphrases for the target category that had considerable overlap across the video and object stimuli.

One 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 stable and discoverable (since both involve atomicity) compared to the connection of unboundedness to substancehood (since both are unspecified for atomicity). Alternatively, we suggest that precisely as a result of the advantage of the bounded category during the Video Session, participants could have come up with more specific conjectures about the class of bounded events that would be hard to abandon when encountering the novel objects. By contrast, participants’ hypotheses about the class of unbounded events might be easier to generalize to a new class of stimuli: as long as a picture lacked structure, neatness, etc., for instance, people would be able to choose it when extending the category of unbounded events. We return to the origin of this effect in later sections.

A possible concern about the results of Experiment 1 is that half of the event pairs in both the training and testing phase used the nature of the affected object as a defining dimension for boundedness: recall that, in those pairs, bounded events involved a single object (e.g., draw a balloon) and their unbounded counterparts involved a mass – usually plural - quantity (e.g., draw circles). Even though the unbounded versions did not involve a substance per se, one might wonder whether this contrast could have facilitated performance in the Picture Session. Answers to the open-ended questions cast doubt on this possibility, since they did not mention a simple object-substance contrast in the videos as the basis for choosing a response in the Picture Session. Nevertheless, a more stringent test of whether events and objects are similar mental entities would eliminate the need to track different types of affected objects during videos. Experiment 2 provided such a test.

3. Experiment 2

In Experiment 2, we replicated Experiment 1 using only the event pairs that contrasted in terms of the nature of the action (e.g., stack a deck of cards vs. shuffle a deck of cards) during the Video Session. Furthermore, we ensured that, among the 10 pairs of videos across the training and testing (Video Session) trials, 5 pairs involved a single canonical object (e.g., fold up vs. wave a handkerchief) and 5 pairs involved a mass quantity (e.g., put up vs. scratch one’s hair). Therefore, the entities in the videos offered no straightforward cues about how to answer test questions during the Picture Session.
3.1. Method

3.1.1. Participants

Forty-eight adults (34 female, 14 male; $M_{\text{age}} = 19.5$, age range: $18.0 - 23.2$) participated in the experiment. All were undergraduates at the University of Delaware and received course credit for participation. The sample size of Experiment 2 was decided by running a power analysis of Experiment 1. With 48 participants, the observed power of the predictor of interest, Condition (Bounded vs. Unbounded) was 0.88.

3.1.2. Stimuli

Only the 10 event pairs that showed the bounded-unbounded contrast in the nature of the action were used from the video stimuli of Experiment 1 (see Table 6 below). The picture stimuli were identical to Experiment 1.

3.1.3. Procedure

Participants were tested in small groups during an online (Zoom) testing session and entered their responses via an online response form. The procedure was identical to Experiment 1 with three exceptions: during training, participants were presented with 4 pairs of bounded and unbounded events (rather than 8 pairs); at test, there were 6 trials (rather than 8) in the Video Session; the surprise testing phase was removed.

3.2. Results

3.2.1. Performance in the testing phase

Results are shown in Fig. 4. Similar to Experiment 1, in the Video Session, participants were better at identifying the Bounded ($M = 78.5\%$, SE = 0.032) than the Unbounded ($M = 66.0\%$, SE = 0.054) category ($\beta = -0.67$, $z = -2.03$, $p = .042$; see Table 7). In the Picture Session, no significant difference was found between the Bounded ($M = 66.7\%$, SE = 0.065) and the Unbounded ($M = 74.3\%$, SE = 0.051) condition ($p > .250$; see Table 8). The average proportion of correct responses in both sessions across both conditions was significantly above chance level ($ps < 0.02$).

We next 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 ($\beta = 1.13$, $z = 2.86$, $p = .004$; Table 9): in the Bounded condition, performance got worse from the Video Session ($M = 78.5\%$, SE = 0.032) to the Picture Session ($M = 66.7\%$, SE = 0.065) ($\beta = -0.67$, $z = -2.37$, $p = .018$), but in the Unbounded condition, performance was not different between the two sessions ($M = 66\%$, SE = 0.054 in Video Session, $M = 74.3\%$, SE = 0.051 in Picture Session, $\beta = 0.46$, $z = 1.66$, $p = .096$).

Lastly, we compared performance in the two experiments. Overall, performance was lower in Experiment 2 ($\beta = -0.50$, $z = -2.33$, $p = .020$; Table 10), presumably due to the shortened training phase. In both experiments, participants were better at identifying the Bounded than the Unbounded category ($\beta = -0.70$, $z = -3.04$, $p = .002$). As already found within each experiment, there was a significant interaction between Condition and Session ($\beta = 1.03$, $z = 3.81$, $p < .001$).

3.2.2. Answers to the open questions

There were 48 answers to the first question (“What kind of videos can get a red frame?”). Similar to Experiment 1, 16 answers (33%) focused on the organization or structure of the stimuli, 6 answers (13%) referred to neatness of the actions, and 3 answers (6%) made reference to goal or purpose. Additionally, 6 responses (13%) referred to hand movements (e.g., “clear movements” for bounded events). Another 13 responses (27%) mainly listed a few examples of the events in the experiment, and 4 responses (8%) did not include any relevant information.

Of the 48 answers to the second question (“What kind of pictures can get a red frame?”), 4 answers (8%) directly addressed the object/substance distinction (e.g., “The pictures with a red frame show whole objects, not liquids”), 11 answers (24%) referred to structure or organization and 5 answers (10%) commented on neatness of the images. Half of the answers (N = 24, 50%) focused on the shapes and lines in the pictures. Only 4 answers (8%) were vague (e.g., “I have no idea”).

3.3. Discussion

Results from Experiment 2 largely replicated what was found in Experiment 1: participants were able to form event categories based on (un)boundedness and to generalize these distinctions appropriately to the object domain. These patterns could not be attributed to possible connections between objects in the videos and the novel entities in the later Picture Session: the design ensured that an equal number of video pairs involved canonical objects vs. mass quantities. Additionally, paraphrases of the features of the target category in participants’ self-generated answers (even though more diffuse than those in Experiment 1) show that observers could extract abstract structural features in both the event and the object domain. These results thus strengthen the conclusion (first

---

7 We reanalyzed performance in the Picture Session with only passers from the Video Session (participants who had at least 4 correct responses out of the 6 trials; 22 out of 24 in the Bounded condition, and 15 out of 24 in the Unbounded condition). During the Picture Session, performance did not differ between the Bounded ($M = 70.5\%$) and the Unbounded ($M = 84.4\%$) condition ($\beta = 0.92$, $z = 1.72$, $p = .085$).
Cognitive Psychology 143 (2023) 101573

Table 6
Video stimuli in Experiment 2. 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).

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

Note: We rotated the videos used in the training (No. 1–4) and the first four pairs in testing phase (No. 5–8).

Fig. 4. Proportion of correct responses in the two sessions of the testing phase in Experiment 2. Error bars represent ± SEM.

Table 7
Fixed effect estimates for multi-level model of event category identification in the Video Session of Experiment 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.05</td>
<td>0.17</td>
<td>6.15***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>−0.67</td>
<td>0.33</td>
<td>−2.03*</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition.

* p < .05, ** p < .01, *** p < .001.

Table 8
Fixed effect estimates for multi-level model of event category identification in the Picture Session of Experiment 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.16</td>
<td>0.27</td>
<td>4.32***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>−0.45</td>
<td>0.49</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition.

* p < .05, ** p < .01, *** p < .001.

posited in Taylor, 1977; cf. Bach, 1986) that event and object representations align, and that notions of atomicity characterize both kinds of representation. Finally, as in Experiment 1, performance got worse from the Video to the Picture Session in the Bounded condition, but no such difference emerged in the Unbounded condition.
4. Experiment 3

Experiments 1 and 2 invited participants to extend their reasoning about events to reason about objects, and found that participants did so in accordance with the (non)atomic properties of the entities involved. In Experiment 3, we directly probed the scope and potency of such cross-domain parallels by training participants to place bounded vs. unbounded events into groups labeled with pictures of objects vs. substances. If abstract underlying construals of (non)atomicity cross-cut event and object cognition, the pattern of labelling bounded events with objects and unbounded events with substances should be natural (and thus easy to adopt) but the opposite pattern should be unnatural (and hence hard or impossible to adopt).

4.1. Method

4.1.1. Participants

Forty-eight adults (32 female, 16 male; \( M_{\text{Age}} = 19.2 \), age range: 18.0 – 22.5) participated in the experiment. All were undergraduates recruited from the University of Pennsylvania and received course credit for participation. The sample size of this experiment was decided based on a pilot study that adopted a similar task. With 48 participants, the observed power of the predictor of interest, Training (Congruent vs. Incongruent) was 0.98.

4.1.2. Stimuli

The video and picture stimuli were identical to Experiment 2. The 10 pairs of videos showing bounded and unbounded events were separated into 2 lists (see Table 11). Each list contained one video from each pair. We counterbalanced whether the event was bounded or unbounded across lists. Each list was further arranged into two presentation orders to rotate videos in the training phase and the testing phase. In each presentation order, 5 bounded events were intermixed with 5 unbounded events (2 bounded and 2 unbounded in the training phase, and the rest in the testing phase).

Each of the 3 object pictures was combined with each of the 3 substance pictures to form 9 unique object-substance pairs; in the experiment, one of these pairs was used twice. The picture pairs were arranged in a pseudorandomized order such that the same picture

### Table 9
Fixed effect estimates for multi-level model of category identification in the testing phase of Experiment 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.12</td>
<td>0.19</td>
<td>5.87***</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>-0.70</td>
<td>0.38</td>
<td>-1.87</td>
</tr>
<tr>
<td>Session (Video vs. Picture)</td>
<td>-0.11</td>
<td>0.20</td>
<td>-0.56</td>
</tr>
<tr>
<td>Condition*Session</td>
<td>1.13</td>
<td>0.40</td>
<td>2.86**</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Condition + Session + Condition: Session.
* \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \).

### Table 10
Fixed effect estimates for multi-level model of category identification in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.14</td>
<td>0.15</td>
<td>7.72***</td>
</tr>
<tr>
<td>Experiment</td>
<td>-0.50</td>
<td>0.21</td>
<td>-2.33*</td>
</tr>
<tr>
<td>Condition (Bounded vs. Unbounded)</td>
<td>-0.70</td>
<td>0.23</td>
<td>-3.04**</td>
</tr>
<tr>
<td>Session (Video vs. Picture)</td>
<td>0.12</td>
<td>0.19</td>
<td>0.63</td>
</tr>
<tr>
<td>Condition*Session</td>
<td>1.03</td>
<td>0.27</td>
<td>3.81***</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + Experiment + Condition + Session + Condition: Session.
* \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \).

### Table 11
Video stimuli in Experiment 3. Each row depicts a pair of events. Participants saw only one event from each pair (as indicated by dashes).

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

A. Papafragou and Y. Ji
would not appear consecutively more than twice. Each member of the pair was then placed on the cover of a folder to create pairs of folders (see Fig. 5; whether a novel object or a novel substance appeared on the top folder was counterbalanced across the trials).

We displayed the picture pairs and the videos as Fig. 5 shows. In each trial, a video appeared on the left of the screen and a pair of folders with pictures appeared on the right.

4.1.3. Procedure

Participants were individually tested online through the PC-Ibex platform (Zehr & Schwarz, 2018). 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 would play next to two folders on the screen. Their task was to decide which folder the video should go into.

The experiment consisted of a training and a testing phase. The training phase had 4 trials and in each trial, participants were informed about which folder the video should go into. Participants were randomly assigned to one of the two conditions. In the Congruent Training condition, the trained video-picture pairing preserved (non-)atomicity across domains: participants were told that videos of bounded events went into the folder that had a picture of a novel object and videos of unbounded event went into the folder that had a picture of a novel substance. In the Incongruent Training condition, the pairing was reversed.

The testing phase had 6 trials; in each trial, participants selected a folder for a new video without receiving any feedback.

4.2. Results

For both Training conditions, a response was coded as correct when participants followed the logic of the training during testing. Specifically, in the Congruent Training condition, a correct response was to choose a picture of a novel object for a video of a bounded event, or a picture of a novel substance for a video of an unbounded event; in the Incongruent Training condition, correct responses paired a novel object with an unbounded event, and a novel substance with a bounded event.

Results are shown in Fig. 6. A mixed-effects model that examined the fixed effect of Training (Congruent vs. Incongruent) did not converge due to close to zero random effects of Subject and of Item. Therefore, we compared performance in the two training conditions using aggregated data (i.e., the proportion of correct responses of each subject). An effect of Training was found ($t(46) = 9.17, p < .001$). In both Training conditions, accurate performance was significantly different from chance level – but in opposite directions: in the Congruent Training condition, people succeeded in pairing a video with a picture that shared the same property of (non-)atomicity 70.8% of the time ($SE = 0.029, t(23) = 7.23, p < .0001$); in the Incongruent Training condition, people disregarded an unnatural match between a bounded event and a novel substance, or between an unbounded event and a novel object ($M = 29.2\%, SE = 0.035, t(23) = -5.93, p < .0001$). In fact, in the overwhelming majority of cases, people in that condition continued to perform ‘natural’ alignments of events and objects. Performance on bounded events did not differ from that on unbounded events in both Congruent ($t(23) = -1.19, p = .247$) and Incongruent ($t(23) = 1.59, p = .125$) Training conditions.
4.3. Discussion

Experiment 3 showed that viewers were able to identify principled parallels between event and object construals: when they received training in placing bounded and unbounded events into groups labelled by objects vs. substances respectively (in accordance with the theoretical notion of atomicity; see Taylor, 1977; Bach, 1986, among others), participants successfully followed the logic of training when sorting further events; but when training violated atomicity, participants failed to follow the logic of the training in their subsequent event-sorting behavior (and in fact went on to sort in ontologically congruent ways). Thus whether people reproduced the trained patterns reflected the perceived organization of otherwise dissimilar stimuli across the event/object divide. The difference between the two training conditions demonstrates constraints on what could be learned, and underscores the fact that participants brought their own preexisting biases about eventhood and objecthood to bear onto the demands of the task.

5. Experiment 4

Currently, it is an open question whether people could spontaneously form a link between bounded events and objects, and between unbounded events and substances even in the absence of overt training. In Experiment 4, we addressed this question by replicating the basic design of Experiment 3 but omitting the training trials. We predicted that, even when viewers freely inspect a single event, we should see evidence of their preexisting, principled ability to link events and objects.

![Fig. 6. Proportion of correct (=training-matching) responses in Experiment 3. Error bars represent ± SEM.](image)

![Fig. 7. Proportion of correct responses in Experiment 4. Error bars represent ± SEM.](image)
5.1. Method

5.1.1. 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 the University of Pennsylvania and received course credit for participation. The sample size of this experiment was decided based on the same pilot study mentioned in Experiment 3.

5.1.2. Stimuli
The video and picture stimuli were identical to Experiment 3. The presentation order of the videos was altered such that the bounded events and the unbounded events were arranged into two blocks rather than being intermixed.

5.1.3. Procedure
The procedure was identical to Experiment 3 with two exceptions. First, there was no training phase: participants selected a folder after watching a video from the first testing trial. Second, participants were randomly assigned to one of two orders. In the Bounded-Unbounded order, the block of bounded events preceded the unbounded one, and this was reversed in the Unbounded-Bounded order. Block order was manipulated because in Experiments 1 and 2 there were differences in how easily bounded vs. unbounded events were mapped onto their counterparts in the object/substance domain.

5.2. 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 that examined the fixed effects of Event Boundedness, Order and their interaction. As shown in Fig. 7 and Table 12, participants were more accurate in choosing a picture of a substance for an unbounded event (M = 72.5%, SE = 0.031) compared to choosing a picture of an object for a bounded event (M = 62.9%, SE = 0.043) (β = 0.44, z = 2.24, p = .025). No effect of order was detected (p > .250), nor was there a significant interaction between the two factors (p = .204). Performance was significantly different from chance level across both event blocks and both orders (ps < 0.025).

5.3. Discussion
Results from Experiment 4 showed that viewers could draw a link between bounded events and objects, and between unbounded events and substances, even in the absence of prior training or specific instructions (cf. Experiment 3). This finding offers strong evidence for the conclusion that the representational units recruited by event and object cognition are similar. As in prior experiments, it was easier to connect unboundedness to substancehood than boundedness to objecthood.\(^8\) We return to the significance of this last finding below.

6. Discussion
Event representations are central to both cognition and language, and yet the precise format of these representations has remained elusive. In this paper, our goal was to throw light onto the unit of event cognition by bridging perceptual-cognitive (see Radvansky & Zacks, 2014; Shipley & Zacks, 2008) and logico-semantic (see Binnick, 2012; Truswell, 2019) accounts of event structure.

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 Kuhn et al., 2021; Wagner & Carey, 2003; Wellwood et al., 2018a, 2018b). 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)

---

Table 12
Fixed effect estimates for multi-level model of picture-video matching in Experiment 4.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.75</td>
<td>0.10</td>
<td>7.62**</td>
</tr>
<tr>
<td>Event Boundedness (Bounded vs. Unbounded)</td>
<td>0.44</td>
<td>0.20</td>
<td>2.24*</td>
</tr>
<tr>
<td>Order (Bounded-Unbounded vs. Unbounded-Bounded)</td>
<td>-0.0004</td>
<td>0.20</td>
<td>-0.002</td>
</tr>
<tr>
<td>Event Boundedness * Order</td>
<td>-0.50</td>
<td>0.39</td>
<td>-1.27</td>
</tr>
</tbody>
</table>

Note. Formula in R: Accuracy ~ 1 + (1|ID) + (1|Item) + EventBoundedness + Order + EventBoundedness:Order.  
* p < .05, ** p < .01, *** p < .001.

---

\(^8\) Overall, participants chose a picture of a novel object 45% of the time, and a picture of a novel substance 55% of the time. Therefore, it was unlikely that the results were driven by a preference for object over substance pictures or vice versa.
phrases resemble mass nouns in that they are unspecified for atomic features. Building on recent evidence that boundedness is a property of\textit{cognitive}, as well as semantic, event representations (Ji & Papafragou, 2020a, 2020b), 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, in conjunction with prior linguistic evidence, this idea could link the unit of both event and object representation in non-linguistic cognition to a single, foundational notion of atomicity.

Our experiments offer direct evidence in support of this idea. In Experiments 1 and 2, 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 3, after explicit training, people learned to organize events in ways that respected atomicity (i.e., grouping bounded events under an object symbol and unbounded events under a substance symbol) but failed to acquire the opposite, atomicity-violating organization patterns. Finally, in Experiment 4, viewers were able to spontaneously draw connections between events and objects without any prior training or explicit instruction. Thus the cognitive representations of bounded/unbounded events and objects/substances align in ways that could be plausibly underscored by a common notion of individuation.

6.1. Implications for theories of event cognition

Our data bear on theories of event cognition. Current cognitive models of event segmentation and event knowledge have not fully explained how events are individuated (e.g., Elman & McRae, 2019; Zacks et al., 2007). According to the segmentation models outlined in the Introduction, event boundaries are transitions from the predictable end of one event to the beginning of the next which is less predictable (Swallow et al., 2009); however, the nature of the boundaries is otherwise unexplained. More broadly, research on event segmentation has not addressed the issue of how people encode the representational unit within event boundaries (i.e., “what happens” during an event). On most such theories (e.g., Zacks et al., 2007), a correspondence between event and object structure is entirely unexpected. By turning the theoretical focus towards the internal temporal texture of individual events as opposed to the transition moments between events, our approach enables us to capture strong and stable intuitions about the ontology of events and objects, and their correspondence. In this sense, the present perspective has a similar motivation to models that move away from studying events in terms of simple placement of boundaries and instead probe the representational content of event construals (including the entities affected as an event unfolds; Altmann & Ekves, 2019). These proposals, however, have not tackled event boundedness and atomicity.

Our current methods were not designed to address the dynamic process of identifying entity units within and across the bounded/unbounded event and object/substance domain. However, recent evidence from online paradigms suggests that (un)boundedness is not only accessible when people explicitly group event tokens into categories but is, in fact, spontaneously encoded during event perception. In Ji and Papafragou (2022), people had to detect a brief visual interruption placed at either the midpoint or endpoint of a bounded or unbounded event. Placement of disruptions affected detection performance for bounded events (especially at endpoints); for unbounded events, the effect was smaller or non-existent, as expected by the fact that these events lack a true mid-or endpoint. These results strongly support the possibility that the systems for perceiving temporal entities are sensitive to the abstract individuation (boundedness) dimensions even when viewers are engaged in an orthogonal task. Other studies likewise suggest that people’s ability to track and quantify over a moving entity is affected by whether the event is an object or a substance (Huntley-Fenner et al., 2002; Pylyshyn & Storm, 1988; Pylyshyn, 2001; vanMarle & Scholl, 2003; vanMarle & Wynn, 2011). It is likely, therefore, that the underlying units of individuation in both the event and object domain offer a powerful way of organizing incoming perceptual information.

What are the mechanisms driving the links between events and objects as mental particulars? Even though our results do not provide detailed evidence, they reveal an interesting asymmetry. Recall that the bounded event category could be identified more easily compared to the unbounded category (see also Ji & Papafragou, 2020a) but the advantage of the bounded event category did not transfer to categorization of spatial entities (Experiments 1 and 2). Moreover, the connection between bounded events and objects was not stronger (and, in fact, tended to be weaker) than that between unbounded events and substances (Experiment 4). There could be several reasons that made it seemingly harder to link a bounded event to a novel object in the current experiments. A bounded event was a structured temporal entity with a salient, self-evident endpoint while a novel object was a structured spatial entity with a salient, easily identified shape and its own function, which was unfamiliar, and unrelated to the event. For this reason, it might have been hard to discern connections between two independent and unrelated structures. By contrast, both unbounded events and novel substances were entities with a largely undifferentiated structure, which could have made it easier to capture their similarities: as long as a picture lacked structure, neatness, etc., for instance, people would be able to relate it to an unbounded event (recall that people’s answers to the open questions typically mentioned lack of structure, messiness etc. in the non-atomic cases regardless of domain).

6.2. Implications for linguistic theories of event structure

Our account directly builds on approaches to event structure within linguistics and philosophy and provides a shared cognitive basis for what is known in this literature as “atomicity” (Bach, 1986; Jackendoff, 1991; Taylor, 1977). Going beyond prior studies that link non-linguistic representations of events and objects to aspectual and count/mass distinctions in linguistic forms (e.g., Kuhn et al., 2021; Wellwood et al., 2018a, 2018b), our data offer a striking piece of non-linguistic evidence for the presence of similarities between spatially extended entities (i.e., objects vs. substances) and temporally extended entities (i.e., bounded vs. unbounded events).

Our data bear on the scope of semantic theories: if similarities between mental properties of events and objects belong (at least in part) to cognitive architecture, then they fall outside the explanatory scope of linguistic theories of lexical aspect and telicity.
Currently, even though most linguistic theories of event structure recognize the potential role of conceptual factors, there is no consensus as to how to apportion the contribution of linguistic vs. extra-linguistic factors to the understanding of individuation across domains (Filip, 2012).

Lastly, compared to segmentation models that often focus on how humans place event boundaries and less on how humans understand events (see previous section), the present approach more naturally connects the content and format of different events to how events are encoded in language. Representing an entity as having or lacking an inherent spatial or temporal boundary can offer a conceptual starting point for talking about an object or an event within and across languages (Perretti et al., 2007; Flecken et al., 2015; Levelt, 1989; Von Stutterheim et al., 2012). The awareness of the atomicity property can also lay the foundation for children to acquire linguistic tools for expressing objecthood and eventhood in their native language (cf. Gleitman, 1990; Pinker, 1989).

6.3. Extensions and future directions

As a first attempt to probe the parallels in the cognitive representation of temporally extended entities (bounded/unbounded events) and spatially extended entities (objects/substances), in our experiments we used a set of videos that underwent a series of norming studies such that the bounded-unbounded distinction was not driven by related event features such as intentionality. Under a similar consideration, we adopted a set of pictures from Li et al. (2009). We acknowledge that the current work does not involve a large number of test items. Follow-up studies may use more stimuli and test a greater variety of entities (e.g., natural events and objects).

Our findings further suggest several directions for future work. First, the current data leave several possibilities unsettled in terms of how the mind comes to conceptualize dynamic stimuli as bounded or unbounded events, and static stimuli as objects or substances. For events, both visual features such as repetition, regularity, and natural breaks (e.g., Kuhn et al., 2021; Strickland et al., 2015) and conceptual features such as functional significance, goals and overall aspects of the context affect how people determine event units (Filip, 2001; Kennedy & Levin, 2008; Mathis & Papafragou, 2022; Zacks & Swallow, 2007). Similarly, visual cues such as shape or non-solidity (e.g., Gentner & Boroditsky, 2001; Soja et al., 1991) and conceptual cues such as the probability of a shape-based function influence objecthood construals (for a combination of both types of cues, see the arbitrariness of structure proposal in Prasada, 1999; Prasada et al., 2002). Pursuing the underlying abstract structure of the two domains (and expanding the stimulus set to include a greater variety of spatial and temporal entities) would strengthen our understanding of how cross-domain links emerge.

Second, our data raise the question of when the observed parallels between event and entity structure arise in the human mind. If – as we have hypothesized - the ability to distinguish between bounded and unbounded events is foundational for the very units of thoughts about events, much like the cognitive ability to distinguish between objects and substances is foundational for the very units of thoughts about entities, cross-domain mappings of the kind reported in this paper should be present early in life. Alternatively, if some of these concepts are not basic, or if the links between the temporal and spatial domain emerge slowly as part of a gradual abstraction over temporal or spatial entity tokens, there is no reason to assume that the patterns reported in the present paper would be replicated in young children. In support of the first possibility, a handful of infant studies have revealed preliminary parallels between object and (bounded) event construal (Sharon & Wynn, 1998; Wood & Spelke, 2005; cf. Maguire et al., 2011).

A final question concerns how conceptual and linguistic notions of atomicity connect. One possibility is that – in accordance with our hypothesis – conceptual atomicity precedes and structures the encoding of atomicity in the verbal and nominal domain. Alternatively, the direction of causation might be the reverse: the conceptual signature of atomicity might arise because of familiarity with the way events and objects are encoded in one’s native language. The first – but not the second - hypothesis predicts that the non-linguistic atomicity of events and objects (and the corresponding parallels) should be similar cross-linguistically. Initial evidence shows that people approach object/substance individuation similarly despite cross-linguistic differences (Li et al., 2009; cf. Imai & Gentner, 1997) but this evidence needs to be expanded to encompass event individuation and their cross-domain correspondence.

6.4. Conclusion

Across four experiments, viewers were found to draw robust and surprising cross-domain mappings between temporal and spatial entities. In concert with prior linguistic theorizing, such mappings point to an ingrained ontological feature that organizes the mental representation of events and objects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data in the manuscript.

Acknowledgements

The authors thank Raevyn Johnson for helping construct the videos, and Daniel Bresnahan, Ariel Mathis and Ugurcan Vurgun for helping collect and/or process the data. This work began while both authors were at the University of Delaware.
Funding
This material is based upon work supported by the Beijing Institute of Technology Research Fund Program for Young Scholars #3240012222302 (Y.J.) and National Science Foundation grant #2041171 (A.P.).

Data availability statement
The data for the present study are available through the Open Science Framework (https://osf.io/c4dax/). “Events and objects are similar cognitive entities” project.

References


