Baltic International Yearbook of Cognition, Logic and Communication

Volume 13 *Events and Objects in Perception, Cognition, and Language* Article 3

2019

Events and Processes in Language and Mind

Alexis Wellwood
University of Southern California

Angela Xiaoxue He
Chinese University of Hong Kong

Haley Farkas
University of Southern California

The Baltic International Yearbook of
Cognition, Logic and Communication

December 2019 Volume 13: *Events and Objects in Perception, Cognition, and Language*

pages 1-25 DOI: http://dx.doi.org/10.4148/1944-3676.1122

ALEXIS WELLWOOD
University of Southern California

ANGELA XIAOXUE HE
Chinese University of Hong Kong

HALEY FARKAS
University of Southern California

EVENTS AND PROCESSES IN LANGUAGE AND MIND

ABSTRACT: Semantic theories predict that the dimension for comparison given a sentence like *A gleebed more than B* depends on what the verb *gleeb* means: if *gleeb* expresses a property of events, the evaluation should proceed by number; if it expresses a property of processes, any of distance, duration, or number should be available. An adequate test of theories like this requires first determining, independently of language, the conditions under which people will understand a novel verb to be true of a series of events or a single ongoing process. We investigate this prior question by studying people's representation of two cues in simple visual scenes: a) whether some happening is interrupted by temporal pauses, and b) whether and how the speed of an object's motion changes. We measured representation by probing people's choice of verb in free-form descriptions of the scenes, and how they segment the scenes for the purposes of counting. We find evidence that both types of cues shape people's representation of simple motions as events or processes, but in different ways.

1. INTRODUCTION

Philosophers and semanticists often assume a tight connection between language and the world: that our words and sentences directly relate to the things and happenings that we want to talk about. Psychologists, in contrast, typically suppose that the language-world connection is inextricably mediated by the mind: that systems of perception and conception shape our experience of things and happenings, and these representations are directly encoded by language. The tripartite connection between language, mind, and the world is important, because there are many ways of representing, and talking about, the same thing or happening. For example, imagine that a cat and a mouse run in a circle: if we think of the cat as agent, the scene is better labeled using the verb *chase*; if we think of the mouse as agent, *flee* is better. This paper is part of a larger project drawing explicit links between language and representation in the visual domain.[1]

Research at the interface between nominal syntax and visual representation has uncovered tight interconnections between labeling and how people view a scene. To get a feel for the influence of language on visual thinking, we might imagine some red confetti-like material sorted into a blue-and-white striped container. If the scene is labeled using a novel noun in count syntax (e.g., *a sib*), people prefer to think that the intended referent is the striped container; in contrast, given mass syntax (e.g., *some sib*), they prefer to think it is the red material (Brown 1957). For the influence of visual representation on language, notice that people prefer to describe a regularly-shaped novel entity using count syntax (e.g., *There is a blicket*), but an irregularly-shaped entity using mass syntax (e.g., *There is blicket*; e.g. Prasada et al. 2002). Such case studies reveal an alignment between the psychologist's 'object' representations (see Rips & Hespos 2015), and the semanticist's 'atomic reference' for count syntax (cf. Link 1983; Gillon 1992, 1999): whatever x counts as N given *a N*, x cannot be arbitrarily divided into more instances of N.

Semantic theory has described parallel commitments for the semantics of verb phrases at least since Vendler (1957), but connections with representation have been relatively underexplored in cognitive psychology (though see Lakusta & Wagner 2016). The distinction on the language side is usually couched in terms of 'telicity' (see e.g. Dowty 1979, Krifka 1989): VPs that include information about a bound, end, or *telos* are called telic predicates (e.g., *run to the park*); VPs that do not include such information are called atelic predicates (e.g., *run in the park*). Telic predicates have been likened to count predicates in requiring that their satisfiers be 'atomic', while atelic predicates (like mass predicates) are neutral in this respect (e.g. Taylor 1977; Bach 1986). Distributionally, telic predicates may be differentiated from atelic predicates in whether they can, for example, describe a single 'happening' measured for its duration (*A ran in the park for an hour, *A ran to the park for an hour*). With respect to atomicity, arbitrary divisions of a single running to the park clearly won't divide into other instances of running to the park, just like a single container generally won't divide into other containers.

Using appropriately constrained scenes, a clear alignment between count syntax and object representation, and between mass syntax and substance representation has been observed (see above). But what about the hypothetical alignment between telic syntax and event representation, and between atelic syntax and process representation? Some experimental evidence already supports a parallel language-to-mind direction of influence in this domain. Given identical scenes involving two objects moving up and down on a screen, a sentence like *A jumped more than B* leads to comparison by number (suggesting event quantification; cf. Barner et al. 2008), but a sentence like *A moved more than B* leads to comparison by duration or number (suggesting process or event quantification; Wellwood et al. 2019). Consistent with these observations, Wellwood et al. (2018b) found that people preferred to label a series of dynamic movements using count syntax if the movement was broken up at non-arbitrary temporal intervals, but to label the same movement with mass syntax if it was broken up at arbitrary temporal intervals.

These distinctions matter. Most importantly for the semanticist's purposes, without an independent grasp on the categories of representation that people deploy when viewing a scene, certain theories will be untestable. For example, Wellwood et al. (2012) suppose that the dimension along which a comparative sentence like *A gleebed more than B* will be understood depends on the ontological category that *gleeb* applies to: if it applies to events (i.e., if *gleeb* is telic), the dimension should be number; if it applies to processes (i.e., if it is atelic) the dimension can be freer. To test the predictions of a semantic theory that depends on ontological distinctions like this, we must have an independent, empirically-motivated grasp of the ontology. In particular, we must be able to say under what conditions a scene is likely to be viewed in terms of a series of events as opposed to an ongoing process.

We take up this issue in the current paper. A better understanding of people's perception of dynamic visual scenes in event or process terms will not only place us in a better position to evaluate semantic theories that depend on this distinction, but it may ultimately help to determine why we should observe the parallels that we do between the nominal and verbal domains. We focus on two features as potential cues that may shape people's perception of motions in highly-constrained visual scenes: the inclusion or absence of arbitrary temporal breaks (cf. Wellwood et al. 2018b,a), and constancy/change of speed (cf. Zacks 2004). And we assess perception in two ways: by evaluating people's choice of predicates (telic vs. atelic) in their descriptions of the motions, and their strategies in segmenting the continuous motion.

As we will show, our experimental results reveal a surprising asymmetry between how people choose to label our simple scenes, and how they count. People's labeling preferences are sensitive to whether or not arbitrary breaks are included (our operationalization of the 'atomicity' feature in the formal semantics literature), while their counting preferences are sensitive to whether or not the object undergoes a speed change (a feature found to be important in the event perception literature). This asymmetry raises new questions about the relationship between linguistic and conceptual encoding, as well as about the relationship between categorization and counting.

1.1. Potential cues: Evidence from semantics and cognitive psychology

What cues in the visual scene could potentially be used to distinguish events and processes? We consider two which strike us as intrinsic to the ontological distinction. The first cue—whether an object's movement is divided into arbitrary temporal slices or not—is invited by consideration of differences between telic and atelic reference. The second cue—whether the velocity of a moving object undergoes a distinct change—can be found in the event representation literature. (In the next section, we will consider another potential candidate, namely whether or not there is a change in the direction of movement; since we take this cue to be extrinsic to the core distinction between event and process, and since the presence/absence of this feature will not distinguish our scenes, it is best addressed in the context of discussing cue validity.)

The first cue—inclusion/absence of arbitrary temporal breaks—is derived from the formal semantics literature, and some initial results linking that literature with properties of visual representation (e.g. Wellwood et al. 2018b). Predicates may or may not refer divisively: atelic predicates are divisive, while telic predicates are non-divisive. The difference may be detected in the intuitive validity of certain simple inference patterns. For example, the VP *push a cart* is a classic atelic predicate, in virtue of its licensing the inference in (1). In contrast, the VP *draw a circle* is telic, in virtue of its failing to license the corresponding inference, (2). In Vendler's (1957) terms, the atelic predicate describes a 'process', any arbitrary part of which is of the same type as the whole, while the telic predicate describes an 'event', arbitrary parts of which are not of the same type. Such data have, thus, been taken to suggest that processes may be divided at arbitrary temporal points and still instantiate that same process (e.g., a division of some moving is still moving), whereas events are different (e.g., a division of a jumping may simply be a rising up, a coming back down, or other non-jumps).

(1) Sally was pushing a cart, but then she stopped.
$\rightarrow$ Sally pushed a cart.

(2) Sally was drawing a circle, but then she stopped.
$\not\rightarrow$ Sally drew a circle.

Wellwood et al. (2018b) offered some initial, indirect evidence for the validity of arbitrary/non-arbitrary division as a cue towards categorization in terms of events versus process. Their experimental participants viewed simple motions that were temporally divided either at arbitrary or non-arbitrary points, and were asked how they would prefer to describe what they saw. They had two sentential options, each containing a novel deverbal noun (cf. Barner et al. 2008): a count syntax option (e.g., *The star did some gleebs*), and a mass syntax option (e.g., *The star did some gleebing*). Following many authors, Wellwood et al assumed that the choice of count syntax would indicate categorization in terms of events. And indeed, they found that trials with non-arbitrary divisions led to a preference for count syntax, while those with arbitrary divisions led to a preference for mass syntax. This evidence counts as indirect for our purposes in two respects: it used noun rather than verb syntax to probe event representation, and it presented participants with a forced choice between two sentences (though see Wellwood et al. 2018a). Stronger and more direct evidence would come from participants' spontaneous descriptions, and in particular, their verb choices. In light of the previous results, however, we hypothesized that the inclusion of arbitrary temporal breaks would make scenes appear more process-like than event-like.

The second cue—changes in speed—is derived from recent evidence in cognitive psychology.[2,3] Zacks (2004) asked participants to segment continuous motions involving two objects, and examined the effect of a series of movement features (e.g., relative position, distance, momentary speed, speed change) on participants' willingness to indicate where meaningful boundaries should be placed within those motions—i.e., event boundaries. Forward-stepwise regression models revealed that speed change is a salient feature reliably predictive of segmentation (and, thus, of meaningful boundaries between events). He manipulated speed change in a continuous manner, varied by pre-defined equations; here, we draw upon speed change to make a potentially categorical difference between events and processes. We hypothesized that: all else equal, a motion lacking any changes in speed would be more likely to be perceived as a process, while a motion with speed changes is likely to be perceived as a series of events—specifically, events whose boundaries align with the points at which the speed changes.

It is worth noting, though, that the possibility of imposing 'extrinsic' bounds on what is intrinsically represented as a process will make the interpretation of segmentation behavior somewhat tricky. To get a feel for this, consider a difference in the interpretive patterns of telic/eventive *jump* and atelic/process *move* combined with measuring vs counting modifiers as in (3) and (4). While (3a) is odd if intended to describe a single occasion of jumping, it is fine if intended to describe a series of jumps, for example those numbering 17, (3b). (4a) is different: it is most naturally interpreted as describing a single ongoing activity. However, ongoing activities like this can be bounded, and counted; this is what is required for (4b). In such cases, what counts as a bound is contextually-determined; the verb's descriptive content doesn't tell us.

(3) a. A jumped for an hour.
    b. A jumped seventeen times.

(4) a. A moved for an hour.
    b. A moved seventeen times.

1.2. Evaluating perception: verb choice & segmentation strategy

How can we assess people's perception of visual scenes? Visual representations are not, of course, directly accessible to us. But some indirect measures can often lend us important insights (see, for example, Wellwood et al. 2015, who measured participants' judgments of the similarity between two events and their reaction times as ways to evaluate event perception). Here we describe two ways that we think will provide some of the best indirect evidence for that perception (cf. Wolff 2003): the choice of main verb used to describe a scene, and where that scene is segmented into countable units.

For verb choice, we draw again on parallels with the mass/count distinction in the nominal domain, which is commonly recruited to detect representation in object versus substance terms.[4] Here in the verbal domain, the relevant linguistic distinction is, again, that between telic and atelic predicates. Telic predicates are importantly 'atomic' and, just like count predicates, provide the right semantic environment to support counting as opposed to measuring (see the previous subsection). Telic verb phrases like *fall asleep* or *jump* are directly acceptable and interpretable with modifiers that imply counting (*three times*) and non-overlap (*over and over*), but not with those that imply measures of temporal duration (*for an hour*), at least not for a single instance of the happening. Atelic predicates like *sleep* and *move* show the opposite patterning.

Thus, a more direct measure than the forced linguistic choice used in previous studies like Wellwood et al's (2018b) is to ask how people choose to describe a scene. In particular, drawing on the language-perception connection observed in the nominal domain, we can ask people to describe what they see, and use their choice of predicate to infer their representation of that scene. To the extent that their verbal descriptions differ in interesting ways, this is a reasonable first step towards inferring properties of their representations of what happened. Specifically, we hypothesized that the selection of telic verbs would reflect representation of a simple dynamic scene in 'event' terms, whereas the selection of atelic verbs would indicate representation in 'process' terms. If the two cues to eventhood proper that we outlined in the previous section are accurate—i.e., inclusion of arbitrary pauses, speed changes in an object's movement—then we would expect to see an increase in telic over atelic language just when those cues are present.

In addition to language choice, how people segment a continuous motion for the purposes of counting should also reflect any cleavage between event and process representation. Events are intrinsically countable units, unlike processes. And indeed, it has been shown that people quantify scenes highly suggestive of 'events' in terms of number, even at a very young age. Wood & Spelke (2005) showed that infants could discriminate large numbers of jumps (i.e., four vs. eight), and cared about the number of jumps even when it was de-confounded with their extent along other continuous dimensions. Moreover, this intrinsic countability might be revealed by tracking things like changes in speed in an otherwise continuous motion (Zacks 2004). Processes, lacking any intrinsic bounds that would support counting, can nonetheless figure in counting behavior; motion that is perceived primarily in 'process' terms can be segmented according to extrinsic boundaries such as a change of location or direction (cf. Magliano et al. 2001). For example, imagine an object moving at a continuous rate back and forth on a screen. Nothing inherent to the movement suggests meaningful units for the purposes of counting, but the boundaries of the screen might provide salient points of segmentation regardless. Therefore, how people segment the scene may provide important cues to how they are representing ongoing motions.

In our experiment, we display simple motions occurring within certain spatial boundaries (i.e., two edges of a computer screen), which may serve as default segmentation points, independently of categorization in terms of event or process. This would manifest in our experiment as counting as 'one' any left-to-right crossing of the screen by the object. Such a segmentation strategy could mean one of two things: an intrinsic categorization in 'event' terms, or one in 'process' terms with an extrinsic bound. Thus, we are particularly interested in deviations from this default. We anticipated, in particular, that people might consider the total motion on a given trial as a single segment (thereby indicating process categorization; with extrinsic spatial boundaries); they might consider a movement across the screen and back as a single segment (thereby indicating event categorization; something like a jump); and of course, on trials with arbitrary pauses, they might consider a movement bounded by those temporal pauses as a unit (thereby indicating process categorization, with extrinsic temporal boundaries).

In short, choices of segmentation points that are different from the default will likely come from people's perception of event boundaries, some of which are intrinsic (reflecting core 'event' categorization) and some of which are extrinsic (built atop core 'process' categorization).[5]

2. EXPERIMENT

In this experiment, we aim to uncover what kinds of visual information will make people think of a happening as either an event or a process.

2.1. Participants

Participants were 60 adults recruited through Amazon's Mechanical Turk. Each were paid $1 for 6 minutes of participation.

2.2. Design

In our experiment, participants saw simple animations and answered two questions for each animation. This section provides an overview of the study design, and the next section provides more details about the features of our animations corresponding to these manipulations.

The basic format of each of our animations consisted of a square object moving from one edge of the computer screen to the other and back, some number of times. Differences between the animations resulted from the manipulation of two factors, corresponding to the two visual cues that may influence categorization in terms of events vs process. The between-subjects factor *Pauses* manipulated to whether or not there were temporal pauses in the object's movement (i.e., Pause, No Pause). The within-subjects factor *Movement Type* (i.e., Constant, Smooth, and Sharp) manipulated whether and how the speed of the object changed during a given animation. The Constant movement type involves no speed changes—thus representing, plausibly, a good candidate for process representation, while both Smooth and Sharp did involve speed changes—thus, plausible candidates for event representation. Smooth and Sharp differ only in whether their change in acceleration was graduated (Smooth) or stepped (Sharp), as we describe in more detail in the next section. We manipulated this visual difference as it might correspond to a difference in the size of the hypothetical events: intuitively, Smooth invites thinking about the composite of a back and forth motion as 'one', while Sharp invites about each single crossing this way.

Following each animation, participants were asked two questions in turn: (i) "what happened?" and (ii) "how many times?" For (i), we chose neutral phrasing that would induce as little bias as possible in the participants' freeform responses. Responses to this question could be coded for many different linguistic features, but we focus specifically on the question of whether their main verb choices were telic or atelic. A response headed by a telic verb signaled, for us, event categorization, and one headed by an atelic verb signaled process categorization. For (ii), the semantics of the question we chose imposes the need for some degree of segmentation of the continuous motion stream into units, while leaving it up to the participant just what should be counted. Ideally, responses to this question would be a count of how many times something matching their description happened. We coded these responses as compatible with one of two segmentation strategies—by the number of screen crossings (henceforth: by crossings) or by the number of back-and-forth movements (i.e., a 2 crossing chunk; henceforth: by back-and-forths).[6]

2.3. *Predictions*

We investigate two possible cues to event versus process representation. We hypothesized (i) that the inclusion of arbitrary temporal breaks would render a scene a better candidate for process over event categorization, and (ii) that sharp changes in speed are more likely to be viewed as indicating meaningful, intrinsic event boundaries. These hypotheses support a couple of clear predictions in light of our study design.

In terms of the description portion of the task, regarding (i) we would expect more uses of atelic verbs (e.g., *move*) in the Pause conditions over the No Pause conditions, and the reverse for uses of telic verbs (e.g., *bounce*). Regarding (ii), we would expect more uses of atelic verbs for the Constant movement type but more uses of telic verbs for the Smooth and Sharp movement types.

In terms of the counting portion of the task, we can first describe some general expectations, and then potential interactions with Pause/No Pause. First, the Smooth movement type features no stepped speed changes, but the graduated change might suggest event boundaries at the start/finish points; if so, we would expect more 'by back-and-forth' segments. For Constant and Sharp, we might expect more 'by crossings' counts, but for different reasons: hypothetically, Constant provides no intrinsic event boundaries, and so segmentation behavior should track extrinsic features like the screen edges; in contrast, Sharp involves speed changes at exactly those points, and so is intrinsically bounded there. At the same time, we expect the Pause condition to make people more likely to view the scene in process terms, yet those pauses may also serve to impose extrinsic temporal boundaries; thus for this condition we might expect (a) more 'by crossing' counts, reflecting greater process categorization; or (b) more counts 'by back-and-forths', reflecting a greater salience to left-edge boundaries. The results regarding this point will thus lend us some insight on the role of extrinsic boundaries on people's representation of the scene, especially when intrinsic boundaries are lacking.

2.4. Stimuli

2.4.1. Movement types

Each animation in our experiment depicted a red square on a light blue background, instantiating one of our three movement types. These are represented schematically in Figure 1. Each schematic marks a single back and forth for each movement type, and highlights whether and where a speed change occurs. The relative distance between the arrows and the size of the shaded region behind those points redundantly indicate the relative speed at which the object moved at a given point in time.

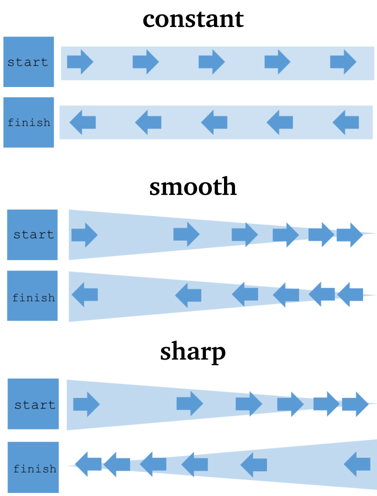

Figure 1: Static depictions of the three movement types. Arrows represent the direction of movement, the width of the shaded band represents relative duration.

The Constant movement type animated the object at a constant speed as it traversed the screen back and forth. Given the constant pace, there was no indication as to event boundaries; if this movement type is to be segmented, extrinsic properties like the screen edges would have to be attended to. The Smooth movement type animated the object speeding up on its left-to-right trajectory, and then slowing down as it returns from right-to-left. We reasoned that the gradual change in velocity at the right edge would not invite perception of an event boundary, but the whole back-and-forth (much like a jumping or a bouncing) might. Hence, the Smooth movement type was designed to elicit events of a larger grain-size than the Sharp movement type, which featured a sharp change in speed at the right edge: the stepped change in speed for the Sharp animations should, we reasoned, lead to greater segmentation at the right edge.

2.4.2. Number of crossings, and pauses

A number of instances of each of our three movement types were programmed, corresponding to different numbers of total crossings, and whether the animation included temporal pauses. For each movement type, animations were included in which the square completed either 6, 10, or 18 crossings (viewed another way: 3, 5, or 9 back-and-forths). Furthermore, each of these animation types came in No Pause and Pause variants.

In the No Pause conditions, the square simply traversed the screen according to the number of crossings programmed for that trial. In the Pause conditions, the square would occasionally pause when it returned to the left edge. The number of times the object paused depended on the number of crossings on that trial. To see how we determined this in detail, consider Figure 2, which displays the number of possible pause locations at the left edge, corresponding to any left-edge appearance of the square except for the first and last in the animation. Thus, the 6 crossing animations made two pause locations available, the 10 crossing animations made four available, and the 18 crossing animations made eight available. In an initial round of programming, we made it so that the square would pause randomly at half of the possible pause locations available given the number of crossings.

However, running the experiment online rendered it infeasible to include every animation corresponding to every combination of n crossings and $n/2$ pauses.[7] Instead, for each animation instantiating n crossings, we selected two $n/2$ Pause variants that the experiment script would randomly select between for a given trial. For the animations with 6 crossings, we included the two possible variants: one where the object pauses only at the first possible pause location, and one where the object only pauses at the second possible pause location. For the animations in which the number of crossings was higher, we chose the animations in which the object paused at points either maximally close to the start/end points, or maximally far from those endpoints. For 10 crossing animations, the pause patterns were either the 1st and 4th pause locations, or the 2nd and 3rd pause locations. For 18 crossings, the object paused either at the 1st, 2nd, 7th, and 8th locations, or at the 3rd, 4th, 5th, and 6th locations.

Because the object's trajectory would be paused only at half of the

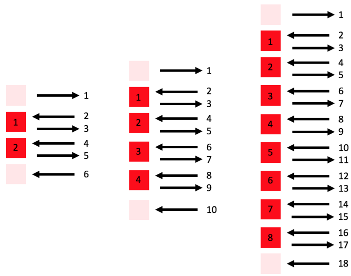

Figure 2: Possible pause locations for 6 crossings (left), 10 crossings (center), and 18 crossings (right). Pink squares mark the start and finish points in the animation. Red squares mark possible pause locations; the number within the square keeps track of the accumulated number of 'back-and-forths'. Each crossing is represented with an arrow, the numbers on the right side of the arrows keeps track of the accumulated number of crossings.

possible pause locations, and because the 'size' of the activity thus divided would be non-uniform, we expected that the inclusion of pauses would appear more as arbitrary pauses in an otherwise continuous activity stream, and thus should be more suggestive of process over event categorization.

2.5. Procedure

The animations for our experiment were programmed using PsychoPy (Peirce et al. 2019). Selected animations, as described in the preceding section, were then screen-recorded using QuickTime. These animations were called upon by an experiment script written in jsPsych (de Leeuw 2015). After an initial instructions screen, participants began one of 9 trials total (3 for each of the movement types, at each of the 3 numbers of total crossings), followed by debriefing screens. The instructions screen indicated the questions they would be asked, and that they would enter their responses into text boxes following each animation; they were furthermore instructed to use their number keys for the second question.

The trial structure can be seen in Figure 3. Each trial consisted of

3 parts: one animation, followed by each of our two questions in turn. Upon completion of the animation, the screen automatically shifted to the first question response screen. This question read, "What happened?", and was paired with a text box. Upon completing their free-form description entered into that box, participants clicked a button to advance the next question screen. This question read, "How many times?", and was paired with another textbox. Following this, participants again clicked a button, which at this point would start a new trial or lead to the debriefing screens.

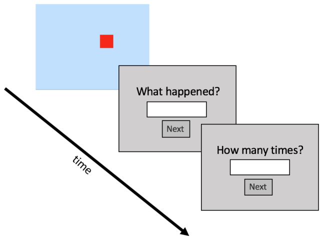

Figure 3: Trial structure for the experiment.

2.6. Results

2.6.1. Coding

All responses were coded and inspected for general patterns by the third author.

For the first question, "What happened?", there were many features of the responses that one could investigate; we focused on the use of telic versus atelic main verbs, and so-classified them according to their infinitival forms. Three verbs accounted for almost 97% of responses: *bounce* (33.7%), *move* (45.3%), and *go* (17.6%). Given the very high proportion of uses of *bounce* and *move*, and given their clear instantiation of opposition in the telic-atelic space,[8] we will focus our analyses on the subset of trials in which these verbs were produced.

For the second question, "How many times?", we were interested in whether the number reported corresponded to the total number of crossings on that trial, or to the number of back-and-forths. For these counts, we considered responses within ±1 to be consistent: e.g., for 18 crossings, we would allow a response of "17" to count as consistent with segmentation based on total crossings. We did this as it seemed reasonable to allow for some small margin of error in participants' counts, or in their recall of those counts, since their responses to this question were recorded after a second, separate screen, and after their responses to the "What happened?" question.

We report the results of our analyses of responses to each of these two questions in turn.

2.6.2. "What happened?"

We first analyzed participants' uses of *move*—a typical process verb, and *bounce*—a typical event verb. Out of 540 trials (60 participants, 9 trials each), *move* was used on 248 trials (45.3%) and *bounce* on 205 trials (33.7%)—these included 13 trials where both verbs were used; on 123 trials, neither was used.

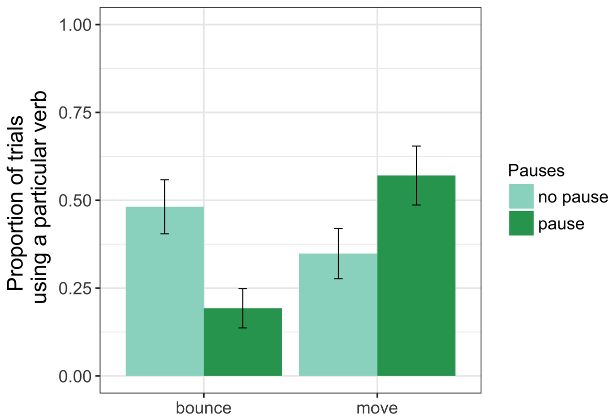

Figure 4: The effect of inclusion of arbitrary temporal pauses on descriptions using *bounce* and *move*.

For each verb, we coded use of that verb as 1 and non-use as 0. Data for *move* and *bounce* were then each entered a mixed-effect logistic regression model, with Pauses (pause vs. no pause) and Movement Type (constant vs. smooth vs. sharp) as fixed effects, and participant and participant-by-movement as random effects. The models indicated a marginally significant effect of Pauses for both *move* (parameter estimate = 3.86, $z = 1.95$, $p = .052$) and *bounce* (parameter estimate $= -5.10$, $z = -2.89$, $p < .01$); but the directions of difference were opposite (see Figure 4)—*move* was used more in the Pause condition (57.0% of all trials) than in the No Pause condition (34.8%), whereas *bounce* was used more in the No Pause condition (48.1%) than in the Pause condition (19.3%).

No effect of Movement was found (all $ps > .20$; Figure 5). These results suggest that the inclusion of arbitrary temporal pauses had a strong effect on verb choices, which reflects people's event perception, over and above the contribution of different patterns of movement—specifically, the inclusion of arbitrary temporal pauses made all movement types more likely to be viewed in process rather than event terms.

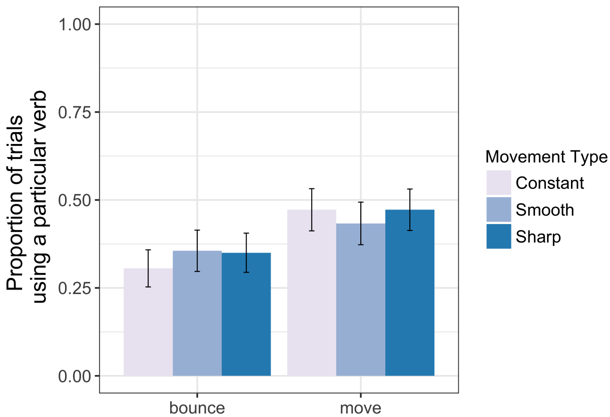

Figure 5: The lack of differentiation between different types of movement patterns on descriptions using *bounce* (telic/eventive)and *move* (atelic/process).

Paired t-tests revealed that in the Pause condition, for all three movement types, *move* was used significantly more than *bounce* (all *ps* < .05); but in the No Pause condition, *move* usage was on a par with *bounce* usage for all movement types (all *ps* > .238; see Figure 6). This further corroborates the above point—that the inclusion of arbitrary temporal pauses makes a scene more likely to be viewed in process terms.

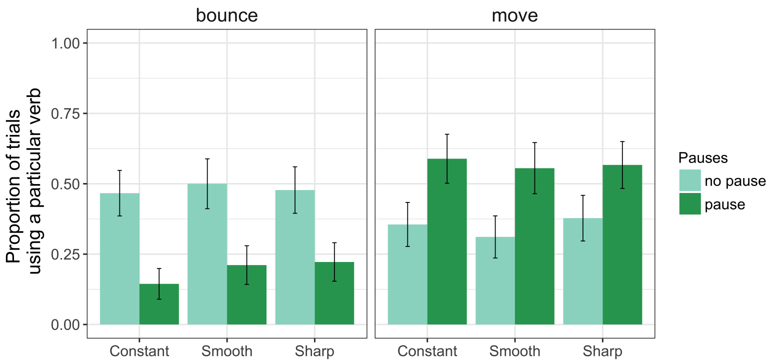

Figure 6: The effect of arbitrary temporal pauses on *bounce* (telic/eventive) and *move* (atelic/process) production is independent of the type of motion depicted.

2.6.3. "How many times?"

We then analyzed participants' segmentation strategy—by the number of screen crossings, and by the number of back-and-forth movements (i.e., half the number of crossings). Out of 540 trials (60 participants, 9 trials each), segmentation was consistent with number of crossings on 250 trials (46.3%) and with number of back-and-forths on 180 trials (14.8%); and on 110 trials neither strategy was used.

For each counting strategy, we coded use of that strategy as 1 and non-use as 0. Data for 'by crossings' and 'by back-and-forths' were then each entered into two mixed-effect logistic regression models, one with Pauses (pause vs. no pause) as the fixed effect and participant as a random effect, the other with Movement Type (constant vs. smooth vs. sharp) as the fixed effect and participant as a random effect. The models indicated a main effect of Pauses for the 'by crossings' strategy (parameter estimate $= -1.59$, $z = -2.66$, $p < .01$) and a marginally significant effect of Pauses for 'by back-and-forths' (intercept parameter $= 1.30$, $z = 1.91$, $p = .057$); but the directions of difference were opposite (see Figure 7)—'by crossings' was used more in the No Pause condition (56.7% of all trials) than in the Pause condition (35.9%), whereas 'by back-and-forths' was used more in the Pause condition (41.5%) than in the No Pause condition (25.2%).

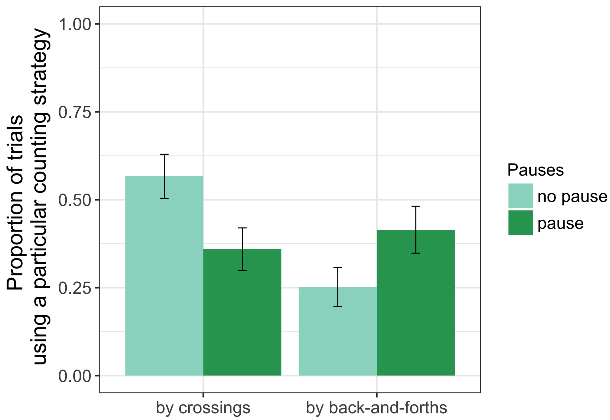

Figure 7: The effect of inclusion of arbitrary temporal pauses on segmentation 'by crossings' versus 'by back-and-forths'.

Our models also indicated main effects of Movement Type. First, the Constant and Smooth movements elicited different uses of counting strategy. Specifically, Constant yielded more 'by crossings' strategy than Smooth (parameter estimate $= 0.83$, $z = 3.01$, $p < .01$), but less 'by back-and-forths' (parameter estimate $= -1.15$, $z = -3.57$, $p < .001$; Figure 8). Second, there was also a difference between Sharp and Smooth: Sharp yielded less 'back-and-forths' strategy than Smooth (parameter estimate $= -0.70$, $z = -2.25$, $p < .01$); it went the opposite direction for 'crossings' but the difference was not significant (parameter estimate $= -.40$, $z = 1.46$, $p = .15$).

These results suggest two points. First, the inclusion of arbitrary

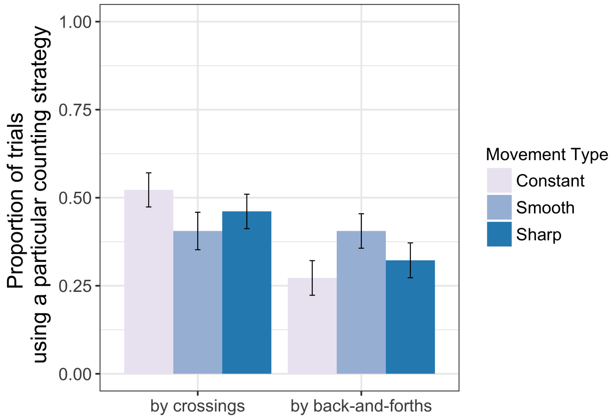

Figure 8: The effect of different movement patterns on 'by crossing' versus 'by back-and-forth' segmentation.

temporal pauses has a strong effect on people's choices of segmentation strategies—specifically, these pauses elicit more 'by back-and-forths' counting and less 'by crossings' counting. Second, the movement type (varying in velocity changes) also affects people's segmentation strategies—specifically, Constant elicits more 'by crossings' counting and less 'by back-and-forths' counting, in comparison to Smooth. Our results are consistent with the prediction that the Constant movement is more reflective of a process and the Smooth movement of events. But the pattern with the Sharp movement is unclear.

Paired t-tests revealed that in the Pause conditions, 'by crossings' and 'by back-and-forths' were used equally for all movement types (all $ps > .32$; see Figure 9). In the No Pause conditions, however, with Constant and Sharp, 'by crossings' was used significantly more than 'by back-and-forths' ($p < .001$ and $p < .01$, respectively), yet there was no difference between them with Smooth ($p = .36$). These results further corroborate the two points made above: that the lack of arbitrary pauses lead to more use of 'crossings' and less of 'back-and-forths', and that Constant and Sharp patterned together, differently from Smooth.

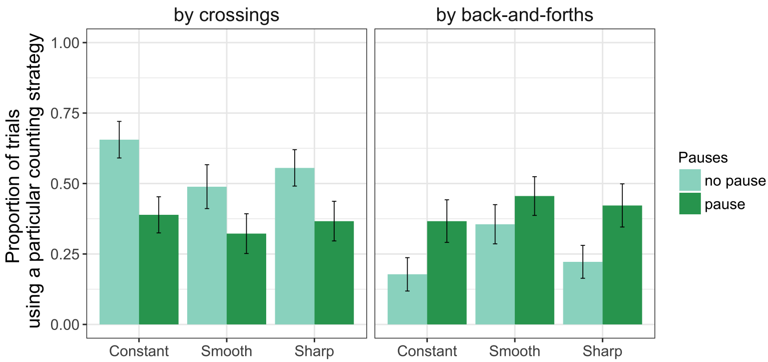

Figure 9: The effect of arbitrary temporal pauses on 'by crossings' versus 'by back-and-forths' segmentation strategies, by movement type.

3. DISCUSSION

We asked: under what conditions do people view a simple motion as a series of events versus an ongoing process. In particular, we considered the role of two potential cues: a) the inclusion/absence of arbitrary temporal breaks, and b) the constancy/change of movement speed. A two-part experiment evaluating participants' verb choices and segmentation strategies reveals the following main findings with respect to these two cues.

For the first cue, the effect of arbitrary temporal breaks is seen on both verb use and segmentation strategy choices. Specifically, people used more atelic verbs in the Pause conditions and more telic verbs in the No Pause conditions, consistent with our prediction. However, people's by-crossing counts decreased in the Pause conditions but by-back-and-forth counts increased, inconsistent with our prediction. We speculated that this is because the points for arbitrary pauses are always at the left edge of the screen, making back-and-forth segments more salient.

For the second cue, the effect of speed change (and thus movement types) is observed only with segmentation strategy, but not with verb choice. The lack of effect on verb use, we speculate, is because in verbal description, people focus more on the abstract structure of the motions (as semanticists have tuned to), paying less attention to concrete differences in speed. People's segmentation strategy choices were largely consistent with our predictions; specifically, for Constant, they counted by crossings, as predicted for a process; for Smooth, they counted by back-and-forths and crossings, a mixture of event and process indications; and for Sharp, they counted by crossings, as predicted for events with boundaries at both the left and right edges.

However, these results also show a surprising asymmetry: the inclusion of arbitrary temporal pauses increases the appropriateness of process categorization (as indicated by a preference for labeling with atelic *move*), but the inclusion of changes in velocity did not decrease this preference (as indicated by the lack of effect of our factor Movement Type on *move*/*bounce* labeling). In a certain respect, these results suggest a disconnect between categorization and counting: labeling, as an indicator of categorization, tracks factors identified in formal semantics (cf. Wellwood et al. 2018b); counting, as an indicator of segmentation, tracks factors identified in cognitive psychology (cf. Zacks 2004). One might have thought that counting was made possible by and depended on one's prior choice of category (cf. Koslicki 1997).

The results also raise a number of questions that should be addressed in follow-up research. For one thing, the present study did not contrast the inclusion of arbitrary versus non-arbitrary pauses, as Wellwood et al. (2018b) did. Based on the previous research, use of the present methodology would lead us to expect to see that the inclusion of non-arbitrary pauses would have the opposite effect (increasing the choice of telic verbs) to that we observed for arbitrary pauses in the present experiment. For another, it may be preferable to eliminate the extrinsic boundaries inherent in our animations (the screen edges) in favor of, for example, circular paths, as this could help decouple any effect of pauses from the effect of location changes. And of course, one will want an account of how the low-level features studied here relate to the higher-order features characteristic of more complex, real-world events (cp. Levine et al. 2017).

Finally, to the extent that we have established firm footing in a real cognitive distinction between events and processes, it would be useful to use materials like those we have devised in an extension of the type of task used by Barner et al. (2008) with the object/substance distinction. They presented participants with novel stimuli that were independently determined to cue object versus substance categorization, labeled the stuff with a novel noun, and showed how the ontological distinction made in perception affected quantification with comparative *more*. In light of the present evidence, we would expect for novel *gleeb more* that the quantification would be more likely to proceed along a continuous dimension, so long as arbitrary pauses are included in whatever movement pattern participants are exposed to. In contrast, it would be more likely to proceed by number when the arbitrary pauses are not included.

ACKNOWLEDGMENTS

This research was supported by an NSF award to AW (BCS-1829225). For helpful discussion that helped to shape the research reported here, we would like to thank Steve Franconeri, Roumi Pancheva, and Toby Mintz, and audiences at New York University, the University of California, Irvine, and the University of Riga, Latvia.

Notes

[1] We occasionally talk of 'perception' rather than 'representation'; for the purposes of the present paper, the two terms are used interchangeably.

[2] This literature does not explicitly consider the distinction of interest here, namely between events and processes as such. Rather, it focuses on the parameters by which ongoing activity is likely to be segmented into a series of discrete events.

[3] Many other factors have been explored with respect to event segmentation, for example the perceived intentions or goals of actors involved in a scene (see Levine et al. 2017 for recent discussion and citations). We are concerned with very simple scenes lacking entities that are likely to be perceived as intentional, deliberately setting aside consideration of such 'higher-order' features.

[4] That is, when people are presented with simple scenes that minimally afford these ontological categories as options for the extension of a novel noun. We do not assume that mass syntax implies the substance category, for example (cf. *furniture*-type nouns). Count syntax, though, does carry semantic commitments towards 'atomicity', or non-divisiveness of reference. We take it that people's representations of concrete objects satisfy such a requirement, but not that count syntax itself implies concreteness, etc.

[5] This characterization corresponds, in the linguistic domain, to analyses in which expressions properly 'about' processes can be embedded into larger expressions 'about' events; see Rothstein 1999, Wellwood 2016, among others.

[6] For example, if the object crossed the screen 8 times, this would correspond to 8 crossings but 4 back-and-forths. Our coding allows for some wiggle room in the counts;

see §2.6.2.

[7]The animations were generated dynamically onsite, but needed to be screen-recorded and uploaded to the internet for the mturk study. Without being selective, we would have had to do this hundreds of times. It seemed to us that the trimmed version described in the text would be sufficient for our purposes.

[8]Roughly following the methodology in Barner et al. 2008, we conducted a norming study on mTurk to determine the telicity status of 213 intransitive verbs (Wellwood et al. 2019). We did this by inquiring about whether a given V, used in a frame like "A V'ed all day", could describe a single happening (durative/process interpretation), or if it must repeat in a series of discrete happenings (iterative/event interpretation). By this measure, *bounce* received an interativity/event score of 93.3%, while *move* received an iterativity/event score of 16.7%.

References

Bach, Emmon. 1986. 'The algebra of events'. *Linguistics and Philosophy* 9, no. 1: 5–16.

Barner, David, Wagner, Laura & Snedeker, Jesse. 2008. 'Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns'. *Cognition* 106: 805–832.

Brown, Roger W. 1957. 'Linguistic determinism and the part of speech'. *The Journal of Abnormal and Social Psychology* 55, no. 1: 1–5.

de Leeuw, J. R. 2015. 'jsPsych: A JavaScript library for creating behavioral experiments in a web browser.' *Behavior Research Methods* 47, no. 1: 1–12.

Dowty, David R. 1979. *Word meaning and Montague grammar*, vol. 7. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gillon, Brendan. 1992. 'Towards a common semantics for English count and mass nouns'. *Linguistics and Philosophy* 15, no. 6: 597–639.

——. 1999. 'The lexical semantics of English mass and count nouns'. In E Viegas (ed.) 'Breadth and depth of semantic lexicons', 51–61. Springer.

Koslicki, Katherin. 1997. 'Isolation and non-arbitrary division: Frege's two criteria for counting'. *Synthese* 112, no. 3: 403–430.

Krifka, Manfred. 1989. 'Nominal reference, temporal constitution and quantification in event semantics'. In Renate Bartsch, Johann van Benthem & Peter van Emb de Boas (eds.) 'Semantics and contextual expression', 75–115. Foris: Dordrecht.

Lakusta, Laura & Wagner, Laura. 2016. 'Conceptualizing the event'. In David Barner & Andrew Scott Baron (eds.) 'Core Knowledge and Conceptual Change', Oxford University Press.

Levine, Dani, Hirsh-Pasek, Kathy, Pace, Amy & Michnick Golinkoff, Roberta. 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* .

Link, Godehard. 1983. 'The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach'. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.) 'Meaning, Use and Interpretation of Language', 302–323. Berlin, Germany: De-Gruyter.

Magliano, J. P., Miller, J. & Zwaan, R. A. 2001. 'Indexing space and time in film understanding'. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 15, no. 5: 533–545.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E. & Lindelø v, J. 2019. 'PsychoPy2: experiments in behavior made easy'. *Behavior Research Methods* .

Prasada, Sandeep, Ferenz, Krag & Haskell, Todd. 2002. 'Conceiving of entities as objects and as stuff'. *Cognition* 83: 141–165.

Rips, Lance J. & Hespos, Susan J. 2015. 'Divisions of the physical world: Concepts of objects and substances'. *Psychological Bulletin* 141, no. 4: 786–811.

Rothstein, Susan. 1999. 'Fine-grained structure in the eventuality domain: the semantics of predicative adjective phrases and *be*'. *Natural Language Semantics* 7, no. 4: 347–420.

Taylor, Barry. 1977. 'Tense and Continuity'. *Linguistics and Philosophy* 1, no. 2: 199–220.

Vendler, Zeno. 1957. 'Verbs and times'. *Philosophical Review* 66: 143–160.

Wellwood, Alexis. 2016. 'States and events for S-level gradable adjectives'. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.) 'Semantics and Linguistic Theory', vol. 26, 166–184. CLC Publications.

Wellwood, Alexis, Hacquard, Valentine & Pancheva, Roumyana. 2012. 'Measuring and comparing individuals and events'. *Journal of Semantics* 29, no. 2: 207–228.

Wellwood, Alexis, He, Angela Xioaxhue, Lidz, Jeffrey & Williams, Alexander. 2015. 'Participant structure in event perception: Towards the acquisition of implicitly 3-place predicates'. *University of Pennsylvania Working Papers in Linguistics* 21, no. 1: 1–9.

Wellwood, Alexis, Hespos, Susan J. & Rips, Lance. 2018a. 'How similar are objects and events?' *Acta Linguistica Academica* 15, no. 2-3: 473–501.

Wellwood, Alexis, Hespos, Susan J & Rips, Lance J. 2018b. 'The object: substance:: event: process analogy'. *Oxford Studies in Experimental Philosophy* 2: 183.

Wellwood, Alexis, He, Angela Xiaoxue & Farkas, Haley. 2019. 'Labeling and quantifying events'. In preparation, University of Southern California.

Wolff, Phillip. 2003. 'Direct causation in the linguistic coding and individuation of causal events'. *Cognition* 88: 1–48.

Wood, Justin N. & Spelke, Elizabeth S. 2005. 'Infants' enumeration of actions: Numerical discrimination and its signature limits'. *Developmental Science* 8: 173–181.

Zacks, Jeffrey M. 2004. 'Using movement and intentions to understand simple events'. *Cognitive Science* 28: 979–1008.