## Title
Temporal Shaping and the Event/Process Distinction



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

**Authors**
Wadle, Douglas
Bansal, Devansh
Wellwood, Alexis C

**Publication Date**
2024

Peer reviewed

# Temporal Shaping and the Event/Process Distinction

**Douglas Wadle (wadle@usc.edu)**
Department of Philosophy, 1445 Jayhawk Blvd
Lawrence, KS 66045 USA

**Devansh Bansal (dbansal@usc.edu)**
Department of Biological Sciences, 3616 Trousdale Parkway
Los Angeles, CA 90089 USA

**Alexis Wellwood (wellwood@usc.edu)**
School of Philosophy, 3709 Trousdale Parkway
Los Angeles, CA 90089 USA

## Abstract

Studies of visual event individuation often consider people's representations of activities involving agents performing complex tasks. Concomitantly, theories of event individuation emphasize predictions about agents' intentions. Studies that have examined simple, non-agential occurrences leave open the possiblity that principles of visual object individuation play a role in visual event individuation. Unearthing principles that may be sufficient for event individuation which are distinct both from predictions about agents' intentions and from visual object individuation, we draw on and extend studies that reveal object and event representation to be deeply analogous in our cognitive economy. We provide evidence that 'temporal shaping' is a sufficient low-level perceptual criterion for the visual individuation of events. In our study, temporal shaping is effected by the introduction of pauses into an otherwise continuous process. Future studies should address other visual mechanisms for introducing temporal shaping (e.g., color changes).

**Keywords:** event individuation, object-event analogy, event semantics, event cognition, mass/count distinction, plurality, boundedness

## Introduction

How do people individuate events in vision? Psychologists studying this question have often focused on people's identification of event boundaries in complex scenes featuring an actor performing a task or series of tasks. Such a focus invites a top-down theory of event individuation with event boundaries that correspond to, e.g., violations of expectation or completion of inferred goals. Although it is acknowledged that the ability to make the relevant predictions must relate to lower-level perceptual features of events, little attention has yet been paid to identifying these features. We rely on and extend studies that reveal object and event representation to be deeply analogous in our cognitive economy – both well-modeled by formal tools in semantics – to provide evidence that 'temporal shaping', i.e. discontinuities in the rate of change of any perceivable feature of the scene, is a sufficient low-level perceptual criterion for visually individuating events. Our results advance our understanding of event representations and how they are deployed in perception and language, providing a clear link between linguistic and perceptual representations and the underlying features of the mind-independent world that they represent.

## Event individuation

A plausible minimal understanding of an event is that it is "a segment of time at a given location that is conceived by an observer to have a beginning and an end" (Zacks & Tversky 2001: 3). Event representations, then, include a starting and endpoint as part of their descriptive content – they are non-arbitrarily temporally bounded – and event individuation will be a matter of picking out salient, non-arbitrary boundaries in occurrences. The study of event individuation in the visual domain has primarily focused on identifying the boundaries of intentional, goal-directed actions – for example, how people segment videos featuring an actor performing some task or tasks (Hard et al. 2011; Ji and Papafragou 2020a, 2020b; Radvansky and Zacks 2011; Zacks and Tversky 2001; Zacks et al. 2007 [etc.]). This tendency, however, introduces a number of complications. We focus on two.

First, relying on scenes involving an actor introduces assumptions of agency and goal-directedness that are not necessary for events, simpliciter; if anything (cf. Steward 2012), goal-directed intentional actions are a proper subset of the set of events. The presence of non-arbitrary temporal boundaries clearly doesn't require agency, or goal-directedness; a branch falling from a tree clearly qualifies. Such examples are acknowledged but quickly set aside in favor of a focus on 'typical events' (Zacks et al. 2007: 273). Furthermore, actions – i.e., goal-directed events performed by an agent – may correspond to a highly specialized category with its own specific principles for recognition and individuation.[1] Any general account of event individuation should address events of the branch-falling sort as well as goal-directed actions.

Second, the scenes are highly complex, and so, plausibly decomposable into multiple levels of lower-level events. For example, in observing someone make a bed, one can segment down into their arranging the top sheet, the pillows, and the bedspread; or into yet further divisions like tucking the sheet under the mattress at each corner, pulling it taut, etc; down to the movement of first the right hand, then the left,

[1] An analogy to face recognition is apt. Faces are plausibly objects, but there are principles governing face recognition that do not apply to broader object recognition (Tanaka and Farah 1993).



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

etc.[2] This significantly complicates the task of finding principles underlying visual event individuation per se because we need to understand the interactions among higher- and lower-level events and to make sure participants are attending to the proper level.[3]

In light of the focus on complex, agent-centered activities, it is no surprise that the principles so far unearthed are relatively high-level: event boundaries are formed when expectations are thwarted, a goal has been achieved, or the agent's direction of movement changes (Radvansky and Zacks 2011; Zacks and Tversky 2001; Zacks et al. 2007). We suggest that a focus on the role of low-level perceptual features in event individuation will be beneficial in overcoming the complications seen in the literature discussed above. Even if the prediction-based view is generally correct, current evidence suggests that the ability to make the relevant predictions requires a basis in low-level perceptual features (Levine et al. 2018). If so, perceptual features indicating goal-directedness, completion, causation, or shifts in intentional action must be isolated before they can reliably so-indicate.[4]

Studying the role of low-level perceptual features in parsing simple, non-agential occurrences can also provide a clearer picture of the perceptual processes involved in event individuation beyond those that may be at work in prediction-based individuation. The object individuation/recognition literature has unearthed many low-level perceptual features that factor in the individuation and recognition of objects – e.g., the role of vertices in determining part-whole relationships (Biederman 1985, 1987). Biederman would likely not have discovered the point about vertices had he restricted himself to manipulations of photographs of everyday scenes. The deep analogy between objects and events suggests that a similar approach to the study of event individuation will be fruitful: we should begin with simpler stimuli than has generally been the case in studies of event segmentation.

To this end, we build on the analogy between the object/substance and event/process distinctions (e.g., Bach 1986) to probe low-level visual features involved in event individuation. To illustrate: Just as two quantities of a substance of type *S* (e.g., mud), taken together, are a single quantity of *S* while two objects of type *O* (e.g., toy), taken together, are not a single *O* (e.g., toy), two bits of processes of type *P* (e.g., walking), taken together, are a single process *P* while two events of type *E* (e.g,. a jump), taken together, are not a single *E*. To represent something as an *E* or *O* as opposed to a *P* or *S* is to represent it as non-cumulative and non-divisive. Such patterning explains why 'count syntax' comfortably applies to object-describing nouns (e.g. *toys*) and event-describing deverbal nouns (e.g., *do a jump*), why 'individuating' adverbials comfortably apply to event-describing verbs (e.g., *jump again and again*), but the reverse isn't true for substance- or process-describing nominals or verbs (e.g., ?*muds*, ?*walk again and again*, ?*do a wander*).

In developing our strategy, like Wellwood et al (2018) and others (discussed below), we rely on a close interaction between linguistic and non-linguistic cognition. In particular, we rely on the properties just discussed concerning event/process verbs and 'individuating' adverbials to probe preferences for different syntactic wrappers for a novel lexical item in descriptions of our scenes. This allows us to test whether participants are parsing the scenes as a series of events or as a continuous process, and to link these preferences to low-level visible features of our scenes. To avoid the influence of agency ascriptions and to isolate a single visible feature of interest, our scenes uniformly involve a simple geometric shape (a circle) that changes, in the given scene, in just one visible feature (position, size, color saturation).

## Prior studies

In one study that gets close to laying bare the low-level visual features involved in event individuation, Wellwood, Hespos, and Rips (2018) examined the influence of 'naturalness' on event versus process classification. In their first experiment, for example, participants watched a star tracing elliptical paths radiating out from a central point. The completed animations, if rendered as an image, would add up to a stylized drawing of a daisy with different numbers of petals. 'Natural' paths traced complete ellipses, broken up by brief pauses at the center point. 'Unnatural' paths involved pauses at other, randomized points. For each animation, Wellwood et al probed subjects' preferences between *The star did some gleebs* (count syntax) or *The star did some gleebing* (mass syntax). Their participants preferred to describe the 'natural' scenes with count syntax (consistent with an event interpretation) and mass syntax with 'unnatural' animations (consistent with event or process.[5] A second experiment showed the same effect for a weaker notion of '(non-)arbitrariness'.[6]

Such a study gives us a point of contrast for a lower-level division of an ongoing flow of activity into discrete events or a merely-interrupted process, illuminating principles that are deployed in individuating events such that they are amenable to count syntax. Nevertheless, it is possible that participants' tracked the paths that Wellwood et al's shapes

[2] It is also worth noting that typical scenes, as in the example of making a bed (taken from Hard et al. 2011), involve the actor acting upon an object. But not all events involve things affected by an agent's activity; consider again the branch falling, or someone jumping, or a cloud drifting by.

[3] Wynn (1996) and Sharon and Wynn (1998) overcome the complications of complex actions in studies of infants' ability to enumerate simple actions and sequences of two simple actions. However, these studies involved a puppet engaged in human-like behaviors, plausibly inviting ascriptions of agency.

[4] Zacks and Tversky (2001, p. 15) suggest that some goals might be evolutionarily transmitted (e.g., the drive to eat), but most will need to be acquired from a combination of social forces and a general-purpose mechanism for causal and goal-based reasoning. Understanding general event individuation will concern the interplay of such a general-purpose mechanism and perceptual learning in the domain of event individuation.

[5] That is, mass syntax comfortably houses lexical items with unindividuated meanings, but it isn't *specified* for non-individuated meanings (cp. *water* and *furniture*; e.g. Gillon 1999).

[6] See Prasada, Ferenz, and Haskell, 2002 for an analogue probing the object/substance distinction.

traversed by relying upon shape-sensitive mechanisms of object individuation. In different blocks of their experiments, subjects were also shown 2D images corresponding to the natural/unnatural or non-arbitrary/arbitrary paths and asked whether they would prefer *There were some gorps* (count) or *There was some gorp* (mass) to describe them. Participants could have used (or felt invited to use) the same (static) shape-based individuating strategies for images and animations, which would be easier to do when the star's starting and stopping point matched for each path segment. Understanding what role, if any, visual object individuation mechanisms play in visual event individuation will be important for contextualizing these results against observations concerning temporal features influencing speech and music segmentation (e.g., pauses, changes in pitch) in the auditory domain.[7]

Still outstanding, then, are principles that may be sufficient for visual event individuation but which cannot plausibly be attributed to object individuation. The obvious place to look for such principles is in the temporal domain. Events unfold in time, objects do not; so, we should look at the temporal shape of events and processes (see Rips & Hespos, 2015 for 'shaping' in the object/substance domain). The temporal shape of an occurrence is a function of the variation in one or more perceivable features of a scene over time. Temporal *shaping* occurs when there are discontinuities in temporal shape – changes in rates of change along any perceivable feature. Our experiments target the role of temporal shaping in event individuation.

## Experiments

We conducted two forced-choice tasks probing preference for a novel verb paired with an individuating or non-individuating adverbial, evaluated against scenes with a temporal pause ('event' scenes) or without ('process' scenes). In Expt. 1, we included low numbers of iterations of a base movement type (1, 2 iterations) or higher (5, 6 iterations), and Expt. 2 included only numbers outside of the subitizable range (6-9 interations). Otherwise, the experiments were identical, and so we discuss them together. We hypothesized that temporal shaping would be sufficient for event individuation, and thus we tested pairs of scenes differing minimally in temporal shape. In particular, we expected that scenes with a temporal gap or pause between iterations of a base animation would be interpreted as multiple occurrences of bounded activity (events), otherwise it would be available to interpret the scene as merely a continuous flow of activity (process). Assuming a tight connection between event individuation and individuating adverbials, we thereby predicted a preference for those adverbials given scenes with temporal gaps.

[7]Any attempt to draw connections across these literatures must contend with the fact that auditory studies treat segments as objects (words, phrases) rather than events; there is no verbal form of these terms ('wording' or 'phrasing') (Chiappe and Schmucker 1997; de Diego-Balaguer et al 2015; Dowling 1973; Johnson et al 2014; Knosche et al 2005; Matzinger et al 2021; Pena et al 2002; Stoffer 1985). Similarly, studies focusing on the influence of segmentation on temporal perception in the visual domain do not directly address the question of the principles we are after. Rather, they assume that a given visual feature (e.g., a pause) is sufficient for segmentation (Liverence and Scholl 2012; Yousif and Scholl 2019).

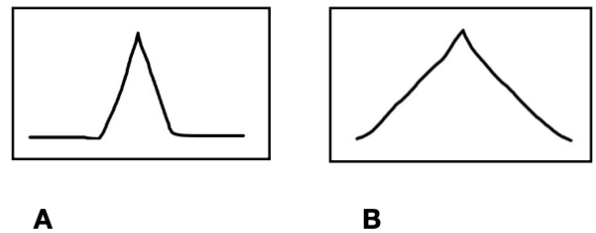


Figure 1: Time plotted against magnitude of change for event (A) and process (B) animations. For any pair of animations, events contained pauses before and after the motion, while processes were continuous across the full duration.

In our design, any cues based on spatial shape that could be extracted from one of an event/process scene pairing are identical to that of the other. Similarly, any 'extrinsic' individuation supported by the boundaries of the animation window is matched, as should any temptation, however remote, to attribute agency to the moving object. For an example, in one of our pairings, a circle begins its movement at the bottom center of the screen, proceeds to the top center (in a straight line parallel to the sides of the window), and returns to its starting position in a smooth motion. In the 'event' version, the circle pauses at the bottom of the screen between each up-and-down movement, whereas in the 'process' version there is no pause, yielding a continuous motion up and down. Any difference detected between members of such a pairing can thereby be attributed to the difference in temporal profile.

We furthermore added generality by looking at kinds of movements that are not path-based (e.g., objects changing in brightness, saturation, and size). Our resulting 8 minimal pairs of scene types were based loosely on the results of a norming study testing known verb iterativity (cf. Barner, Wagner, and Snedeker 2008), with one element of the pair rated highly iterative and the other the opposite. Iterativity, we assume, with others, presupposes events.

### Participants

We recruited 60 participants for each of Expts. 1 and 2 (120 participants in total) on Amazon's Mechanical Turk (Mturk) using Cloud Research (Hauser et al. 2022). All participants self-reported to be native speakers of English, and participation was restricted to accounts located in the United States. No participants were excluded. Expt. 1 took approximately 10 minutes to complete, and participants were compensated \$2. Expt. 2 took approximately 15 minutes, and participants were compensated \$3.

### Methods

**Visual stimuli.** Using PsychoPy (Pierce, J. et al., 2019) and JavaFX, we programmed minimal pairs of animations with each instantiating a simple temporal opposition within a base time $t$ of 2 seconds (Figure 1 schematizes this event-process distinction in $t$). For event scenes, $t$ consisted of two periods of rest (1/4 $t$, each) with the movement contained in the middle half. For processes, $t$ consisted of continuous change along the same dimension. A blue circle was used as the sole event participant throughout. Figure 2 depicts 3 of the 8 scene types we developed, alongside intuitive glosses consisting of iterative/non-iterative verbs that loosely correspond to the movements they depict. We programmed tokens in each event/process variant for each scene type in 1-9 iterations.

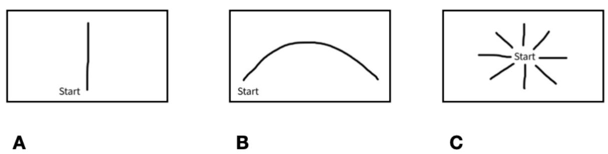


Figure 2: Schematic representations of three scene types: **A** cs02 (*jump*, *run*), **B** cs03 (*hop*, *move*), **C** cs05 (*pop*, *grow*).

**Linguistic stimuli.** Participants were offered two possible descriptions of each animation, with variations in: the novel verb (unique for each scene type); whether the adverbial implies individuation (yes for (1-a) and (2-a)); and whether the adverbial was appropriate for single (1) vs multiple iterations of the base animation (2). The sample options in (1) with the novel verb *prot*[8] were presented for cs01 (*bounce*, *float*) when instantiated in a single iterations ($n = 1$) and those in (2) when instantiated in multiple iterations ($n > 1$).

(1) a. The circle protted once. +IND
b. The circle protted for a little bit. −IND

(2) a. The circle protted every second or so. +IND
b. The circle protted for awhile. −IND

**Procedure.** Participants accepted the HIT on Mturk and were directed to a Firebase-hosted experiment page. Following an instructions screen, participants were presented with a series of 32 trials with the following structure: a fixation cross, followed by an animated scene (the visual stimuli), and then a screen showing the prompt, "Which sentence would you prefer to describe that animation?" alongside the appropriate selections instantiating the options just described (the linguistic stimuli). Participants pressed 'f' to select the individuating option and 'j' to select the non-individuating option. Two lists of animations were used, each of which presented either the event or the process version of each scene type at some number of iterations, with 30 participants seeing each list in each experiment. No participant saw both the event and process variant of the same scene type. In each experiment, the trial order was fully randomized, and there were no catch trials. In Expt. 1, animations were tested in 1, 2, 5, and 6 iterations, and in Expt. 2, animations were tested in 6-9 iterations.

## Results

**Experiment 1.** We found that participants chose the individuating adverbial option (*once*, *every second or so*) at about 67% for event scenes and about 59% for process (Figure 3). A Fisher's exact test revealed a significant effect of animation type on the choice of individuating over non-individuating syntax ($p < 0.001$), indicating a stable relationship between satisfaction of the individuation requirement (i.e., inference to non-arbitrary 'shaping') and pauses in what was otherwise the same flow of activity. For process scenes, ambivalence between the choices likely reflects a combination of (i) the lack of hard-coding of non-individuation for adverbials like *for a little bit* or *for awhile* (cp. Gillon 1999 for mass syntax), and (ii) the possibility of using 'extrinsic' boundaries (e.g., the circle reaching the screen's edge) to individuate otherwise perceptually-continuous scenes.

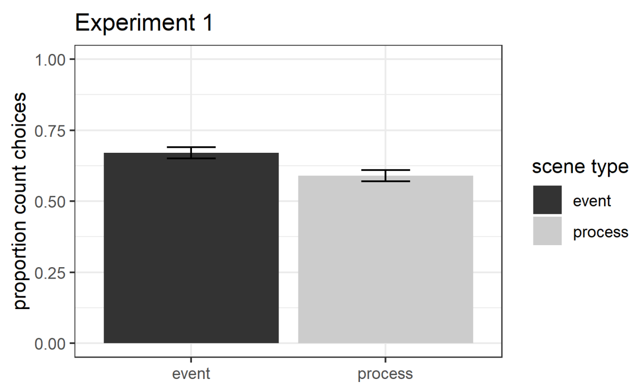


Figure 3: Greater preference for an individuating adverbial given 'event' as opposed to 'process' scenes in Expt. 1.

We did not expect that, overall, people would prefer individuating adverbials for both our event and process scenes. Suspecting that this pattern could be illuminated by looking at the specific numbers of iterations we tested, we dug into the effect of scene type at each iteration (Figure 4). Here, we observed a marked increase in individuating-adverbial choices at low numbers of iterations (1,2), but a general decline and then flattening out of that preference as that number increased (5,6). Plausibly, these differences were driven by the fact that, given only one iteration of the displayed activity, our intended difference in temporal shape was not available to participants; the only difference between the event/process variants of each scene type in 1 iteration is that the object moves more quickly in 'event' than in 'process'. However, no comparison be-

[8]Our novel verbs varied in their final consonants: event scenes had novel verbs ending in stops (e.g., *to prot*) and process scenes ended in fricatives (e.g., *to prosh*). In designing the novel verbs, we worked to ensure that the phonetic shape of the novel verb provided no inconsistent cue to boundedness (see the tentative findings of Kuhn, Geraci, Schlenker, and Strickland, 2021). Planned studies will probe the extent to which such a cue would affect our results.

tween these was available in our between-subjects manipulation of scene type. That participants' preferences began to diverge at $n = 2$ is suggestive in this regard. However, the still-elevated preference for the individuating option at 2 iterations implicates the possibility of a general preference for 'extrinsic' individuation when the number of iterations is in the subitizing range, which has not previously been observed.

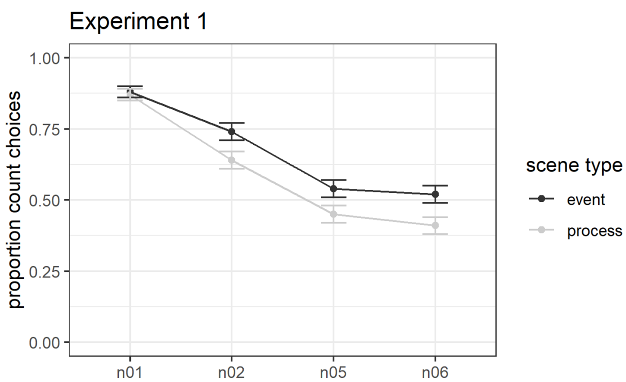


Figure 4: In Expt. 1, the event/process difference was neutralized given only one iteration of the activity, only evening out (and supporting differential preference) at higher numbers.

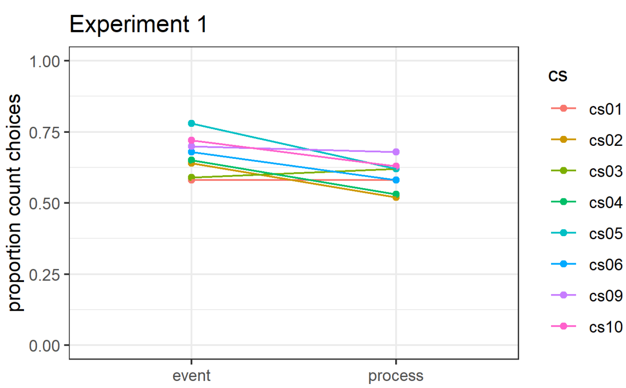


Figure 5: Proportion of individuating selections by scene type in Expt. 1 for each minimal pair tested.

Additionally, considering the proportion of individuating choices for event versus process animations within each scene type, we observed some pairings which saw less of a difference than others (Figure 5)[9]. That is, while cs02 (*jump*, *run*), cs04 (*swing*, *move*), cs05 (*pop*, *grow*), cs06 (*blink*, *shrink*), and cs10 (*vanish*, *diminish*) clearly showed the predicted difference, cs01 (*bounce*, *float*), cs03 (*hop*, *move*), and cs09 (*flash*, *glow*) did not. There are many potential explanations available here based on confounding factors that could have invited greater attention to 'extrinsic' bounds, or obscured the salience between a temporal pause and some other change; these will need to be probed in future work.[10]

Expt. 2 addressed the question of whether numbers of iterations in the subitizing range were responsible for the overall heightened individuating selections for process scenes.

**Experiment 2.** Testing only animations instantiating 6-9 iterations of each scene type and its event/process variants, we observed a greater divergence in the selection of individuating adverbials than we saw in Expt. 1. Participants in Expt. 2 chose the individuating option at about 62% for event scenes and 47% for process (Figure 6). Once again, a Fisher's exact test revealed a significant effect of animation type on the selection of individuating adverbials ($p < 0.0001$). According to our hypothesis, scenes with a regular pause should provide a sufficient cue to individuate and hence selection of an adverbial that semantically selects for individuation; animations without such pauses can, but need not, support such selection (e.g., through 'extrinsic' individuation based on the boundaries of the animation). Unlike in Expt. 2, we also observed a consistent preference for average selection of the individuating option across numbers of iterations (Figure 7).[11]

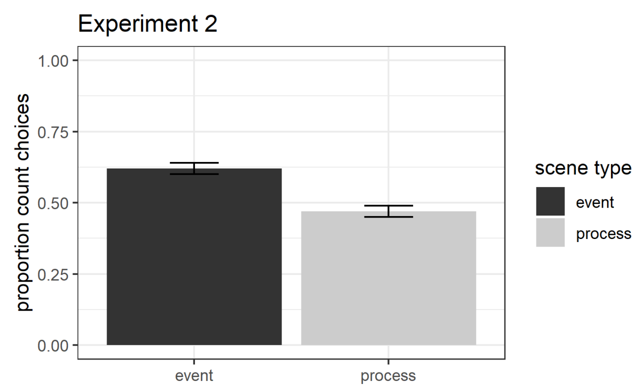


Figure 6: The event/process difference in the selection of individuating adverbials was magnified in Expt. 2.

In Expt. 2, we also observed a greater asymmetry in the preference for sentences with individuating adverbials across all candidate sets (Figure 8). The exception was cs10 (*vanish*, *diminish*). It is possible that after-images played some role in obscuring the intended temporal shaping in those scenes.[12]

[9]Pilot studies tested two further scene types, cs07 and cs08, which involved both path and size changes. The present studies focus on scene types in which only 1 dimension changed over time.

[10] The terminal points of the paths of cs01 and cs03 may have been particularly salient given standing associations with the sort of path traversed. Applying intuitive physics to cs01 could make the lowest point of the path particularly salient, and the back and forth arcing motion of cs03 replicates that of a metronome. For cs09, after-images – particularly at lower iterations – might have played a role in obscuring the distinction in temporal shape.

[11]The figure is suggestive of an increase in individuating selections from at least the lowest number of iterations tested to the highest. We used a Fisher's exact test to compare these two subsets of the data (responses to $n = 6$ and $n = 9$) within the process animations; the difference was not statistically significant ($p = 0.099$).

[12]See fn. 10 for a parallel suggestion about cs09 in Expt. 1.

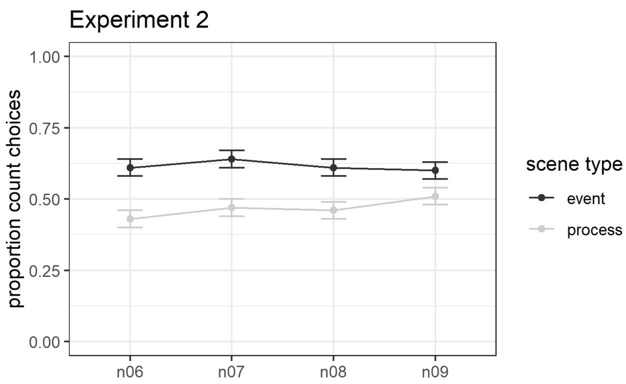


Figure 7: In Expt. 2, the event/process difference was smooth across higher numbers of iterations.

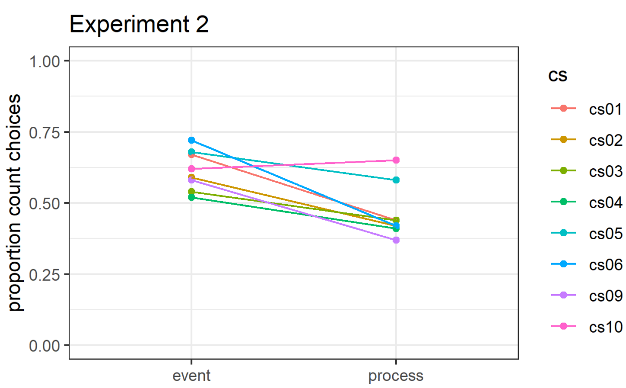


Figure 8: Proportion of individuating selections by scene type in Expt. 2 for each minimal pair tested.

## Discussion

Our results suggest that a difference in temporal shaping – effected, in the present study, by introducing pauses between repetitions of a basic activity – is sufficient for generating an event boundary, and, thereby, for treating the observed activity as a bounded, countable unit. In the absence of such temporal shaping, the activity is seen as an unbounded process. That the inclusion of pauses was the only relevant difference between our event and process pairings allows us to rule out the possibility that one could individuate simply by imagining an object tracing a static shape, thereby recruiting object individuating mechanisms for event individuation. First, the same paths were used in both event and process versions of our path-based animations. If path shape drove event individuation, we should have observed similar results across both the event and process variants of each scene type, but we did not. Second, we used non–path-based animations (e.g., animations displaying changes in hue or saturation over time) with similar results. Hue and (especially) saturation changes are not inputs to general object recognition algorithms.

These results suggest that there are distinct principles for object and event individuation based on low-level perceptual features – more cautiously, that there are principles that are sufficient for the individuation of events from low-level perceivable qualities that are independent of any such principles for object individuation.

Interestingly, differences in responses to our path-based and size-based scenes show that change of direction is not sufficient for event individuation. If it were, we should have seen equivalent selections of the sentences with individuating adverbials in our path- and size-based process scenes (those without intervening pauses) as we did in their event counterparts (those with intervening pauses), since the total path traversed by the circle included a screen crossing and a return in both types of animations. Similarly for scenes involving changes in size – with the object either expanding towards the edges of the animation window and then shrinking back to original size, or else shrinking towards the center of the animation window and then returning to original size. This suggests either that temporal shaping by introduction of a pause is sufficient for event individuation or that it is the conjunction of pause and the change in direction that together provide sufficient temporal shaping for event individuation. We are presently designing animations to test whether pauses, alone, are sufficient.

The use of a simple geometric shape in our animations should minimize the tendency to ascribe agency. While there are, of course, compelling demonstrations that people are willing to ascribe agent-like properties to geometric shapes (see Heider & Simmel, 1944, for the classic demonstration), we think it is highly unlikely that our adult participants were construing our scenes in this way. At any rate, if the blue circle were construed as or like an agent, its inferrable 'goals' would presumably be the same for each of our scene types, regardless of their event/process variations. As such, our results support the conclusion that, in relevantly non-agential contexts, temporal shaping (of at least one low-level perceptual feature) is indeed sufficient for the individuation of events.

The present study focuses on the use of a pause to provide temporal shaping. The positive results, here, suggest that further work should be done to examine the impact of temporal shaping of other visible features on visual event individuation. Our results suggests that (regular) change of direction, size, and saturation are not sufficient to individuate events, but they say nothing about other possible visible features that could be shaped temporally. Further studies should also look at the interaction of temporal shaping across multiple features in more complex scenes and, ultimately, in agency-involving scenes. Suitable care should be taken in designing visual stimuli to avoid a potential role for spatial shaping (which could implicate object individuation mechanisms). Furthermore, it is unknown whether the animation window itself can serve as an extrinsic frame for individuating a given activity. Here too, designing visual stimuli that avoids such a potential confound will not be easy. Color suggests itself as an obvious dimension for avoiding these potential difficulties.

## Acknowledgments

This project was funded in part by a National Science Foundation grant (NSF BCS-1829225) to PI Alexis Wellwood. Our deepest thanks go to Anaiya Raisinghani, Araceli Carmona, and Mohammad Naqvi for helping to design, implement, and test variations on potential event/process animations that were used in pilot studies and in the studies presented here.


## References


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

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, *32*, 29-73.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115-147.

Chiappe, P., & Schmuckler, M. A. (1977). Phrasing influences the recognition of melodies. *Pyshonomic Bulletin and Review*, *4*, 254-259.

de Diego-Balaguer, R., Rodriguez-Fornells, A., & Bachoud-Levi, A. C. (2015). Prosodic cues enhance rule learning by changing speech segmentation mechanisms. *Frontiers in Psychology*, *6*, 1476.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perpetual Psychophysics*, *14*, 37-40.

Gillon, B. (1999). The lexical semantics of English mass and count nouns. In E. Viegas (Ed.), *Breadth and depth of semantic lexicons* (p. 51-61). Springer.

Hard, G., B. M.; Recchia, & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology*, *140*(4), 586-604.

Hauser, D., Moss, A., & Rosenzweig, C. (2022). Evaluating cloudresearch's approved group as a solution for problematic data quality on mturk. *Behav Res*.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259.

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

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

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month olds. *PLoS ONE*, *9*, e83546.

Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). Perception of phrase structure in music. *Human Brain Mapping*, *24*, 259-273.

Kuhn, J., Geraci, C., Schlenker, P., & Strickland, B. (2021). Boundaries in space and time: Iconic biases across modalities. *Cognition*, *210*.

Levine, D., Buchsbum, D., Hirch-Pasek, K., & Golinkoff, R. M. (2018). Finding events in a continuous world: A developmental account. *Developmental Psychobiology*, 1-14.

Liverence, B., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 549-54.

Matzinger, T., Ritt, N., & Fitch, W. T. (2021). The influence of different prosodic cues on word segmentation. *Frontiers in Psychology*, *12*, 622042.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., ... Lindeløv, J. (2019). Psychopy2: experiments in behavior made easy. *Behavior Research Methods*.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604-607.

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

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *WIREs Cognitive Science*, *2*, 608-620.

Rips, L. J., & Hespos, S. J. (2015). Division of the physical world: Concepts of objects and substances. *Psychological Bulletin*, *141*(4), 786-811.

Sharon, T., & Wynn, K. (1998). Individuation of actions from continuous motion. *Psychological Science*, *9*, 357-362.

Steward, H. (2012). Actions as processes. *Philosophical Perspectives*, *26*(1), 373-388.

Stoffer, T. H. (1985). Representations of phrase structure in the perception of music. *Music Perception*, *3*, 191-220.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225-245.

Wellwood, S. J., A.; Hespos, & Rips, L. J. (2018). The object : Substance :: Event : Process analogy. *Oxford Studies in Experimental Philosophy*, *2*, 183-212.

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, *7*, 164-169.

Yousif, S. R., & Scholl, B. J. (2019). The one-is-more illusion: sets of discrete objects appear less extended than equivalent continuous entities in both space and time. *Cognition*, *185*, 121-130.

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

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