Event Segmentation

Jeffrey M. Zacks and Khena M. Swallow

Washington University in Saint Louis

ABSTRACT—*One way to understand something is to break it up into parts. New research indicates that segmenting ongoing activity into meaningful events is a core component of perception and that this has consequences for memory and learning. Behavioral and neuroimaging data suggest that event segmentation is automatic and that people spontaneously segment activity into hierarchically organized parts and subparts. This segmentation depends on the bottom-up processing of sensory features such as movement and on the top-down processing of conceptual features such as actors' goals. How people segment activity affects what they remember later; as a result, those who identify appropriate event boundaries during perception tend to remember more and to learn more proficiently.*

KEYWORDS—*event perception; segmentation; motion; intentions*

Look at the scene depicted in Figure 1a. Though most viewers will eventually figure out what is shown, many will have an easier time understanding the alternative version shown in Figure 1b. What is the difference? The first picture fractures the scene in a way that obscures its natural part structure, whereas the second respects that structure. For quite a while, psychologists have known that in order to recognize or understand an object people often segment it into its spatial parts (e.g., Biederman, 1987). A new body of research has shown that just as segmenting in *space* is important for understanding objects, segmenting in *time* is important for understanding events.

Event segmentation is the process by which people parse a continuous stream of activity into meaningful events. Recent developments in perceptual psychology and cognitive neuroscience have provided new insights into the role of event segmentation in human cognition. In this article, we review three. First, event segmentation appears to be an automatic, ongoing component of human perception. Second, segmentation during perception scaffolds later memory and learning. Third, specialized neural mechanisms identify event boundaries by tracking significant changes in physical and social features.

Address correspondence to Jeffrey M. Zacks, Department of Psychology, Washington University, Saint Louis MO 63130; e-mail: jzacks@artsci.wustl.edu.

SEGMENTATION IS AUTOMATIC

Much of the research on event segmentation has used variants of a procedure developed by Newtson (1976), in which participants watch a movie of some activity and press a button whenever, in their judgment, one meaningful event ends and another event begins. This task produces event-boundary judgments that are reliable across viewers and within viewers across time (Newtson, 1976; Speer, Swallow, & Zacks, 2003). The boundaries they identify tend to be readily nameable "chunks," corresponding to subgoals an actor is attempting to satisfy in order to fulfill the larger goal of the activity. For example, boundaries in "putting up a tent" might be placed when the tent is staked out and when the rain fly is attached (see Fig. 2). Event boundaries are hierarchically structured, such that fine-grained events are clustered into larger coarse-grained events. For example, observers segmenting a movie of a woman making a bed tend to identify events such as removing individual pillowcases and also to identify the removal of all the pillowcases as a larger event (Zacks, Tversky, & Iyer, 2001). The consistency and structure of these results suggest that the Newtson procedure taps into ongoing naturally occurring perceptual processing. But there is a problem: The task requires that observers attend to event boundaries and make decisions about where they occur. Such task demands may change the nature of the perceptual processing involved.

Stronger evidence that event segmentation is naturally ongoing comes from indirect measures, particularly functional neuroimaging studies. In one experiment (Zacks, Braver, et al., 2001), participants viewed a series of movies of everyday activities (e.g., washing dishes, fertilizing a houseplant) while changes in brain activity were recorded with functional magnetic resonance imaging. After first passively viewing the movies, participants segmented the same movies twice, to identify event boundaries at two temporal grains. *Fine* boundaries marked the smallest events the participants found natural and meaningful, and *coarse* boundaries marked the largest events they found natural and meaningful. These boundaries were then used as markers to analyze the brain-activity data from the initial passive-viewing session. During passive viewing, regions in the posterior and frontal cortex showed transient increases in activity that began several seconds before each event boundary and peaked several seconds after the boundary. Responses were larger for coarse boundaries than for fine ones. Because the
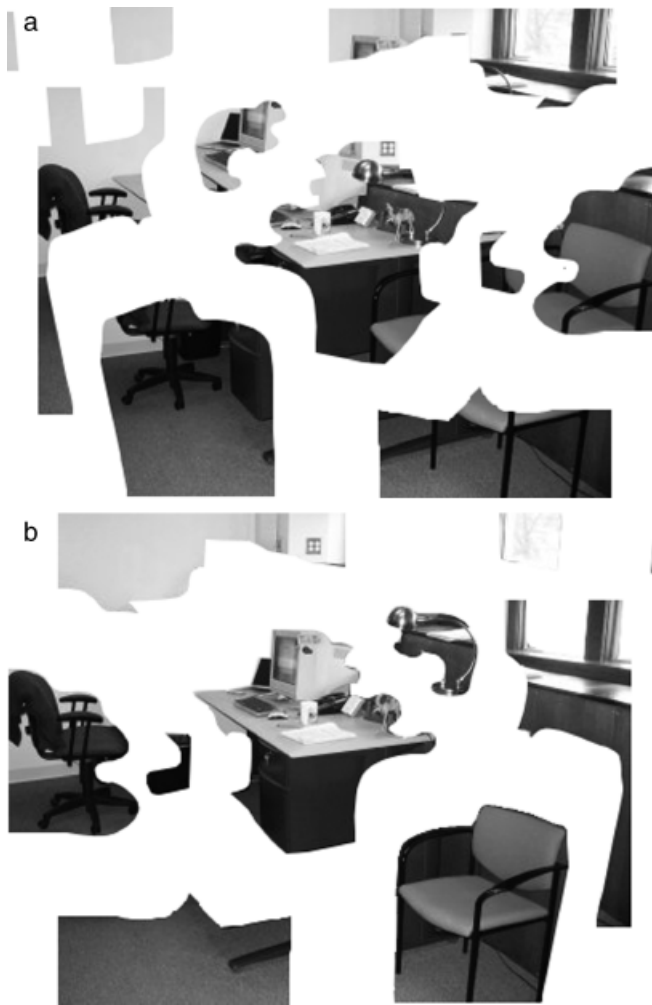
Fig. 1. Two versions of the same photograph, broken up in different ways. The first (a) is harder to identify as an office, because it is broken up in a way that obscures the scene's natural part structure. In the second version (b), the natural parts are preserved.

critical brain data were acquired before participants learned of the segmentation procedure, these changes cannot be attributed to overt or covert performance of a laboratory-specific task. These results strongly imply that brain processes correlated with event segmentation are a normal part of ongoing perception.

The preceding data make a compelling case that event segmentation is automatic. However, segmentation still may be affected by observers' attention and goals. Indeed, there is evidence that observers can adapt their performance of the button-pressing segmentation task based on situational needs. For example, observers adjust the temporal grain of their segmentation based on explicit instructions, the sort of information they are trying to learn from a stimulus, and how much they know about the activity they are watching (see Zacks & Tversky, 2001, for a review). An important question for future research is whether these variations in overt task-related behavior reflect changes in ongoing perception or changes in the decision processes that are specific to the button-pressing task.

SEGMENTATION GUIDES MEMORY AND LEARNING

One important consequence of perceptual segmentation is that the resulting segments can form the basis of memory and learning. New data indicate that, when it comes to remembering the details of what has recently happened, those individuals who are better able to segment an activity into events are better able to remember it later (Zacks, Speer, Vettel, & Jacoby, 2006). In this study, older adults segmented movies of everyday events (e.g., setting up a tent, planting a flower bed). Each person's segmentation was compared to the segmentations of the other observers to determine whether he or she identified event boundaries that were similar to those of the others or placed event boundaries in idiosyncratic locations. Although there is no gold standard for establishing whether one has segmented a movie correctly, observers generally agree on the locations of event boundaries. So, if a particular observer's segmentation deviates from this norm, chances are something is amiss. All participants later completed a test requiring them to discriminate still pictures taken from the movies from similar pictures that were not from the movies. Those participants who segmented "well" showed better memory for the visual contents. Importantly, this strong relationship was observed above and beyond the effects of individual differences in cognitive ability and the presence of senile dementia.

Segmenting an activity well is not simply a matter of identifying the right event boundaries; it also requires tracking how sets of fine-grained events group together into larger meaningful units. Recent studies suggest such grouping is important for learning new activities (Hard, Lozano, & Tversky, in press; Lozano, Hard, & Tversky, in press). In these studies, participants watched movies demonstrating simple assembly tasks and segmented them at both a fine and coarse grain. They then were asked to perform the assembly task they had just watched. The experimenters analyzed the degree to which participants grouped events into hierarchical units. A number of experimental manipulations affected the degree of hierarchical segmentation, and in all cases segmenting hierarchically was associated with better performance of the learned task. Together, these data suggest that event boundaries form anchors for long-term memory and that interventions that encourage people to identify appropriate event boundaries can improve memory for what has happened and the learning of new skills.

NEURAL AND INFORMATION-PROCESSING MECHANISMS

We have argued that event segmentation is an automatic component of normal perception that shapes how people remember and learn. How does the brain perform this segmentation? Evidence indicates that the brain and mind track features of one's environment and that when a salient feature changes unpredictably an event boundary is perceived (Zacks, Speer, Swallow,
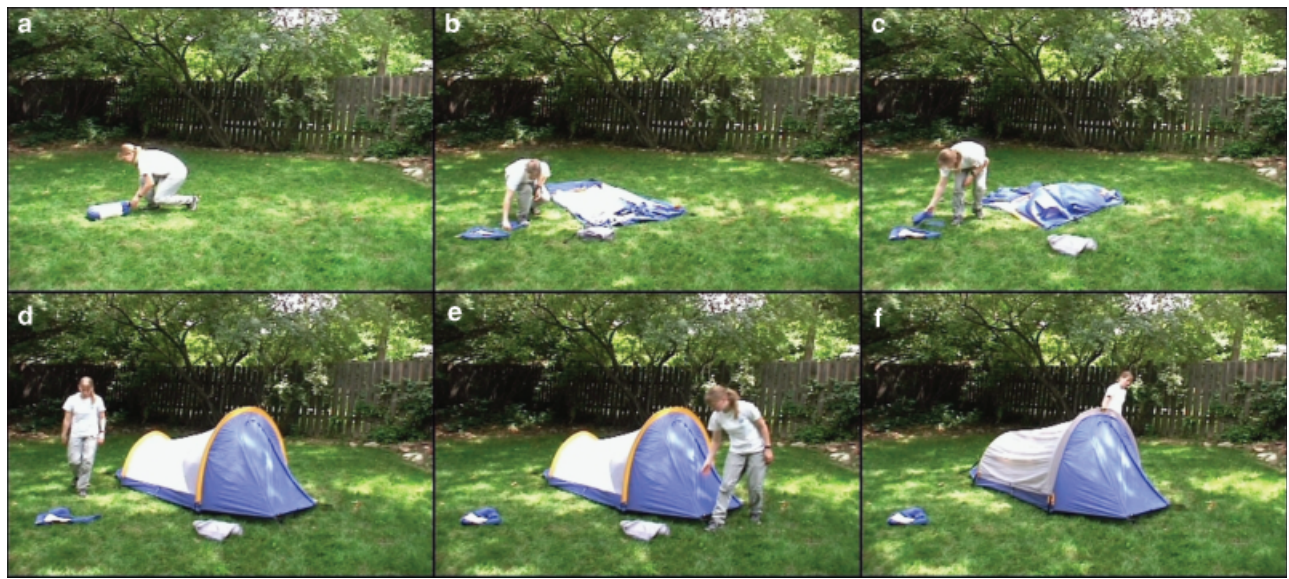
Fig. 2. Example of event boundaries. These frames from a movie of a woman pitching a tent show the six coarse-grained event boundaries selected most frequently by a group of younger and older adults (Zacks, Speer, Vettel, & Jacoby, 2006, experiment 2). These boundaries marked the ends of events that could be described as (a) put down the tent, (b) spread it out, (c) insert the front tent pole, (d) stake out the ends of the tent, (e) stake out the sides, and (f) attach the rain fly.

Braver, & Reynolds, in press). The critical features may include sensory features, such as color, sound, and movement, and conceptual features, such as cause-and-effect interactions and actors' goals. Sensory features likely are processed in a primarily bottom-up fashion, in which the nature of the processing is determined primarily by perceptual input. Processing conceptual features, however, likely relies on top-down processing that integrates an observer's representation of the current event with previously stored knowledge. For example, segmenting events based on an actor's goals requires maintaining a representation of those goals over time and often will depend on prior knowledge about the actor's dispositions and abilities.

Sensory Features: The Movement of Objects and People
One hint about the sensory features that are important for event segmentation came from the neuroimaging study described previously (Zacks, Braver, et al., 2001). In that study, the most active sites of transient brain responses at event boundaries appeared to be in visual-processing areas known to process movement information. These areas collectively are called the human MT complex (denoted MT+, because they are thought to be homologs of areas in the medial temporal cortex of the macaque monkey that are known to be motion-sensitive). A follow-up to the Zacks, Braver, et al. (2001) study identified the MT+ in individual observers and confirmed that the areas activated by event boundaries did include this area (Speer et al., 2003), suggesting that people use movement cues to identify those boundaries. These results motivated recent experiments exploring the quantitative relationship between movement features and event segmentation.

In one set of experiments (Zacks, 2004), participants viewed simple animations depicting two objects moving around the screen and segmented them to identify fine or coarse event boundaries. The animations had been generated either by people who had been asked to play a video game in which one player controlled each object and tried to achieve some goal (e.g., chase the other object and catch it), or by a random algorithm that produced animations with velocities and accelerations that were matched to those of the human-generated animations. Because the animations consisted only of the movements of two objects, they could be completely characterized, quantitatively, in terms of movement features such as distance, speed, and acceleration. In all experimental conditions, participants tended to segment at points when movement features changed—for example, when objects were accelerating quickly or when they reached a point of being maximally close to each other and turned away. (See also Hard et al., in press). Thus, movement changes are strongly related to event segmentation.

If the processing of movement information in the MT+ contributes to the segmentation of activity into events, then one should expect this brain area to track both movement information and event-boundary locations when people view animations such as those in the Zacks (2004) study. A recent neuroimaging study indicates that this is the case (Zacks, Swallow, Vettel, & McAvoy, 2006). In this experiment, participants viewed simple animations while brain activity was recorded. Separate scans were used to identify the MT+ in individual observers. Activity in the MT+ increased as the velocity of the objects increased and also at event boundaries. Thus, the processing of movement information appears to be well situated to play a causal role in the detection of event boundaries.

Conceptual Features: Actors' Goals

Although movement features account for a substantial part of where people segment activity, they are not the whole story. Movement is more strongly tied to segmentation when viewers identify fine-grained units or segment random animations than when they identify coarse-grained units or segment animations depicting intentional activity (Zacks, 2004). This suggests that people depend on other sources of information, such as inferences about actors' intentions and goals, to understand the larger structure of activity. One piece of direct evidence for the importance of goals in event segmentation comes from infant perception. Baldwin, Baird, Saylor, and Clark (2001) familiarized infants with one of two brief movies depicting a woman cleaning a kitchen. Each movie depicted a goal-directed action sequence (replacing a fallen dishtowel or storing an ice cream container in the freezer). After the familiarization phase, infants were shown excerpts from the films with 1-second pauses inserted into them. The pauses were placed either at the moment when the woman achieved a salient goal (e.g., grasping the towel) or shortly before. The infants looked longer at the excerpts when the pauses were placed before the goal completions, suggesting that they found those more disruptive.

Another piece of evidence for the importance of goals comes from a neuroimaging study of reading events in texts (Speer, Reynolds, & Zacks, in press). In this experiment participants' brain activity was recorded while they read a narrative text describing the activities of a small boy. After the imaging session, each participant segmented the text into events. Prior to the experiment, trained raters coded the narrative to identify points of conceptual change, including changes in actors' spatial location and goals. Event boundaries in the narrative were associated with brief increases in brain activity that were similar in timing and location to those for live-action movies. Many of these areas also responded to changes in the narrative features—and the brain responses to event boundaries were entirely mediated by the effects of the narrative features. This suggests that both physical-movement features (changes in location) and changes in actors' goals play important roles in the segmentation of activity into events.

These studies begin to provide the data base for a mechanistic account of how observers segment ongoing activity into events. However, the available data afford only the barest outlines of such an account. We regard the detailed characterization of the relation between bottom-up and top-down processing in event segmentation as one important goal for future research. Further, we believe that a number of little-studied features, from purely sensory to purely conceptual, must be important for event segmentation. Toward the sensory end are features such as sound, lighting, and contact between actors and objects. Toward the conceptual end are features such as goals and social conventions. In the middle are features such as sequential statistical structure—that is, the order in which events tend to occur. The systematic exploration of these bases for segmentation is a second important research goal.

IN THE COURSE OF EVENTS

The previous sections have reviewed recent evidence supporting three conclusions about event perception. First, event segmentation is an automatic component of ongoing perceptual processing. Second, how people segment activity in real time has significant effects on how they remember it later: Events form the units of memory encoding, so identifying the right events leads to good memory and learning, and identifying the wrong events leads to poor memory and learning. Third, there are specialized neural systems that process features including movement and goals in order to use changes in those features to identify event boundaries. These findings have implications for education and for clinical practice. For education, they suggest that interventions that help people appropriately segment events will help them remember and learn from those events. For clinical practice, these findings suggest that some cognitive deficits may reflect impaired event segmentation. A small number of studies indicate that event segmentation is impaired in patients with damage to the prefrontal cortex, in those with schizophrenia, and in those with mild Alzheimer-type dementia (see Zacks et al., in press, for a review). The fact that event segmentation is related to later memory (as discussed previously) raises the possibility that it may be possible to remedy some memory deficits by improving segmentation.

Segmentation is a powerful perceptual operation. By reducing a continuous flux of activity to a modest number of discrete events, a perceiver can achieve terrific economy of representation for perception and later memory. Segmentation is not only economical; it also allows one to think about events as discrete parts that can combine in new ways. This sort of cognition is notoriously difficult with continuous, unsegmented representations. For this reason, people generally perceive space as consisting not of continuous gradations of color and texture but of spatially coherent objects. The same holds in time: Just as much as our everyday perceptual world is made up of discrete objects, it is made up of discrete events.

Recommended Reading

Baldwin, D.A., & Baird, J.A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), *Early social cognition* (pp. 215–240). Hillsdale, NJ: Erlbaum.

Newtson, D. (1976). (See References)

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., et al. (2001). (See References)

Zacks, J.M., Speer, N.K., Swallow, K.M., & Braver, T.S. (in press). (See References)

Zacks, J.M., & Tversky, B. (2001). (See References)

REFERENCES

Baldwin, D.A., Baird, J.A., Saylor, M.M., & Clark, M.A. (2001). Infants parse dynamic action. *Child Development*, *72*, 708–717.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–117.

Hard, B.M., Lozano, S.C., & Tversky, B. (in press). Hierarchical encoding of behavior: Translating perception into action. *Journal of Experimental Psychology: General*.

Lozano, S.C., Hard, B.M., & Tversky, B. (in press). Perspective-taking promotes action understanding and learning. *Journal of Experimental Psychology: Human Perception & Performance*.

Newtson, D. (1976). Foundations of attribution: The perception of ongoing behavior. In J.H. Harvey, W.J. Ickes, & R.F. Kidd (Eds.), *New directions in attribution research* (pp. 223–248). Hillsdale, New Jersey: Erlbaum.

Speer, N.K., Reynolds, J.R., & Zacks, J.M. (in press). Human brain activity time-locked to narrative event boundaries. *Psychological Science*.

Speer, N.K., Swallow, K.M., & Zacks, J.M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective & Behavioral Neuroscience*, *3*, 335–345.

Zacks, J.M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*, 979–1008.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655.

Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., & Reynolds, J.R. (in press). Event perception: A mind/brain perspective. *Psychological Bulletin*.

Zacks, J.M., Speer, N.K., Vettel, J.M., & Jacoby, L.L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology & Aging*, *21*, 466–482.

Zacks, J.M., Swallow, K.M., Vettel, J.M., & McAvoy, M.P. (2006). Visual movement and the neural correlates of event perception. *Brain Research*, *1076*, 150–162.

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

Zacks, J.M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*, 29–58.