Cognition 112 (2009) 201–216

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Using movement and intentions to understand human activity

Jeffrey M. Zacks *, Shawn Kumar, Richard A. Abrams, Ritesh Mehta

Washington University, Psychology Department, Campus Box 1125, St. Louis, MO 63130-4899, United States

ARTICLE INFO

Article history:
Received 9 February 2008
Revised 11 March 2009
Accepted 18 March 2009

Keyword:
Event perception

ABSTRACT

During perception, people segment continuous activity into discrete events. They do so in part by monitoring changes in features of an ongoing activity. Characterizing these features is important for theories of event perception and may be helpful for designing information systems. The three experiments reported here asked whether the body movements of an actor predict when viewers will perceive event boundaries. Body movements were recorded using a magnetic motion tracking system and compared with viewers' segmentation of his activity into events. Changes in movement features were strongly associated with segmentation. This was more true for fine-grained than for coarse-grained boundaries, and was strengthened when the stimulus displays were reduced from live-action movies to simplified animations. These results suggest that movement variables play an important role in the process of segmenting activity into meaningful events, and that the influence of movement on segmentation depends on the availability of other information sources.

1. Introduction

Event segmentation is the process by which people break up a continuous, fluid activity into meaningful events. For example, an observer of a baseball game might perceive it as consisting of innings, at-bats, and individual pitches. For activities less structured than baseball there may not be strong norms for where the boundaries between events go; nonetheless observers often show excellent agreement in placing their boundaries (Dickman, 1963; Newtson, 1976). Neuroimaging and EEG studies suggest that event segmentation is an ongoing concomitant of normal perception—people do it all the time, whether or not they are consciously attending to events and their boundaries (Sharp, Lee, & Donaldson, 2007; Speer, Reynolds, & Zacks, 2007; Zacks, Swallow, Vettel, & McAvoy, 2006c; Zacks et al., 2001a).

The event boundaries that people identify are important for later memory. In immediate memory, one's representation of the current event appears to act as a working memory buffer, with information in the buffer more accessible than comparable information from previous events (Gernsbacher, 1990; Speer & Zacks, 2005; Swallow, Zacks, & Abrams, 2009). After viewing a movie, pictures taken from event boundaries are remembered better than pictures taking from intervening moments (Newtson & Engquist, 1976). Asking viewers to attend to events at different temporal grains affects their later memory (Hanson & Hirst, 1989, 1991; Lassiter, 1988; Lassiter & Slaw, 1991; Lassiter, Stone, & Rogers, 1988). Finally, across individuals, event segmentation is correlated with later memory for events (Zacks, Speer, Vettel, & Jacoby, 2006a). This suggests that better understanding of event segmentation may be important for understanding and improving memory. Understanding event segmentation also may be helpful for constructing systems to automatically segment continuous data streams such as video recordings or sensor data (Mann & Jepson, 2002; Rubin & Richards, 1985; Rui & Anandan, 2000)—if one can quantitatively characterize the cognitively natural breaks between events it may be possible to identify them automatically and use them to select key frames for visualization or units of analysis.

* Corresponding author. Tel.: +1 314 935 8454; fax: +1 314 935 7588.
E-mail address: jzacks@artsci.wustl.edu (J.M. Zacks).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2009.03.007

How does the mind-brain identify event boundaries from the continuous stream of sensory input? *Event Segmentation Theory* (*EST*) (Zacks, Speer, Swallow, Braver, & Reynolds, 2007) proposes that event segmentation arises as a side effect of ongoing understanding. To understand an ongoing event, an observer processes incoming information to generate predictions about what will happen in the near future. Such predictions allow for adaptive proactive actions, and are a key feature of models of control in psychology (Neisser, 1967) and neuroscience (Schultz & Dickinson, 2000). EST proposes that everyday activity includes substantial sequential dependency, which can help prediction. For example, consider watching a friend make salad. One can make predictions about what will come next based on conceptual features such as inferred goals—if the friend takes out a knife this implies the goal of cutting something. One can also make predictions based on perceptual features such as those that arise from biological motion—if the friend begins chopping with a particular frequency and amplitude those parameters are likely to be stable. According to EST, perceivers take advantage of such predictability by maintaining working memory representations of the current event, called *event models*. However, when one event ends and another begins (when the friend finishes the salad), many of the predictive relationships will no longer hold. At such points one's predictions will tend to generate more errors, and it would be adaptive to update one's event models to capture the new event that has begun. EST proposes that when prediction error increases transiently, comprehenders update their event models. This is perceived as an event boundary. Event boundaries are processed simultaneously on multiple timescales, a suggestion supported by physiological studies (Sharp et al., 2007; Speer et al., 2007; Zacks et al., 2001a). For identifying *fine-grained* event boundaries, the system monitors prediction error over shorter time intervals and identifies brief transient increases; for identifying *coarse-grained* event boundaries, the system monitors longer intervals and identifies increases that are larger and longer.

EST and other psychological accounts of event segmentation (Newtson, 1976; Zacks et al., 2007) argue that segmentation depends on the processing of feature changes—particularly those that are not predicted. Feature changes may be conceptual, such as changes in actor's goals, or perceptual, such as changes in movement patterns. Previous studies have provided evidence that both sorts of feature changes are correlated with event segmentation. Conceptual changes predict the locations of event boundaries when comprehenders read or hear narratives (Zacks, Speer, & Reynolds, 2009), and when they view movies (Zacks, Swallow, Speer, & Maley, 2006b). Physical changes—particularly movement—have been studied both qualitatively and quantitatively. One qualitative study coded the positions of actors' bodies at 1-s intervals, and found that changes in body configuration were associated with event segmentation (Newtson, Engquist, & Bois, 1977). Another study (Hard, Tversky, & Lang, 2006) used a simple animation based on the classic event perception work of Heider and Simmel (1944). The animation was coded for qualitative changes in motion, such as changes in direction or speed of motion. Such changes were correlated with event segmentation.

A pair of previous studies used simple animations of pairs of point objects moving on a white background (Zacks, 2004; Zacks et al., 2006c). Movements were characterized quantitatively by computing the speed and acceleration of each of the objects, the distance between the objects, and their speed and acceleration relative to each other. Viewers' segmentation of such animations was significantly correlated with changes in these movement variables (Zacks, 2004). Correlations were stronger when participants segmented the activity into fine-grained events and less strong when they identified coarse-grained events. Correlations were stronger for stimuli that viewers interpreted as depicting random motion rather than goal-directed actions. Brain activity in regions specialized for motion processing covaried both with changes in movement information and with changes in objects' speed (Zacks et al., 2006c). This is consistent with the hypothesis that comprehenders perceive event boundaries in part due to processing changes in movement variables. However, the simple stimuli used in these experiments place limits on the conclusions they can support. Naturalistic everyday action provides rich cues from facial expression, eye gaze, and the objects in the environment. One possibility is that observers may monitor movement information when there is little else available, but for rich depictions of naturalistic activity other cues dominate movement.

This psychological approach is consistent with work in artificial intelligence on the individuation of actions from motion. Thibadeau (1986) described a computational scheme for identifying event boundaries in simple animations. The animations are coded to provide descriptions of the state of the on-screen world for each frame. Changes in states of the system correspond to changes in, for example, the position of an object. Second-order changes are changes in first-order changes. Event boundaries are drawn from the set of second-order changes. Thus, constant-velocity motion does not constitute a boundary, but acceleration does. A formal analysis provided by Rubin (1985) came to a similar conclusion. They showed there is a class of motion transitions that can be reliably identified from the two-dimensional projection of a three-dimensional motion sequence and that corresponds to psychological boundaries. The primitive transitions are starts, stops, and discontinuities of force. These can be composed to form 15 motion transitions (e.g., a stop and a start can be composed into a pause). Rubin and Richards proposed that such motion transitions may correspond with observers' perceptual segmentation of motion sequences. Force discontinuities can be directly identified from object's acceleration, provided that mass is constant. Thus, this analysis converges with Thibadeau's (1986) hypothesis that second-order changes are important for detecting psychological boundaries. Subsequent work (e.g., Mann & Jepson, 2002) has built on such formal analyses to design systems that can segment motion sequences in video.

Thus, previous results suggest that when viewers watch everyday activities, they perceive event boundaries in part due to processing changes in the movement in those activities. However, important questions remain. First and foremost, to this point there has been no quantitative evidence that movement features predict how viewers segment naturalistic action. It is tempting to generalize from the studies just reviewed that used simple animations. However, such animations are quite impoverished compared to live-action movies of naturalistic action. Comprehension of such animations depends solely on rigid body motion information. On the other hand, everyday activity is rich with information about the objects being acted upon, the expressions and eye gaze of actors, and the nonrigid articulation of the body. It would not be surprising if the presence of this additional information rendered the relationship between movement information and event segmentation negligible. Therefore, a first important question is this: Are movement variables robustly correlated with event segmentation when viewing naturalistic everyday activities? The two experiments reported here asked this question by recording movements while an actor performed a set of everyday activities and then asking a set of viewers to segment those activities.

A second, related question is: If movement variables are correlated with event segmentation during naturalistic activities, does removing some of the other information that live-action video provides change this relation? If segmentation depends in part on information about objects, gaze, and facial expression, then removing those cues might strengthen the dependence of segmentation on movement features, and perhaps to change the nature of the relations between movement and segmentation. Experiments 2a and 2b investigated this possibility by comparing segmentation of live-action videos to segmentation of simple animations generated from the movement information captured by the motion tracking system.

Third, if movement and segmentation are related when viewing naturalistic action, does this relation vary with segmentation grain? Events on different timescales may be characterized by changes in different sorts of features, ranging from physical perceptual-motor features for events on the timescale of seconds to abstract conceptual features for events on the timescale of months to years (Barker & Wright, 1954). Everyday events occupy timescales from a few seconds to tens of minutes, and there is evidence that within this range of timescales different features characterize events at different levels. More fine-grained events (with median lengths of 10–15 s) are more strongly associated with specific actions on objects whereas more coarse-grained events (with median lengths of 40–60 s) are more associated with action contexts (Zacks, Tversky, & Iyer, 2001b) or conceptual features such as goals and causes (Baldwin & Baird, 1999). Such results support an interpretation that for everyday activities fine-grained events are more perceptually determined, whereas coarse-grained events are more conceptually determined. Data from simple animations also are consistent with this view, indicating that movement variables are more strongly related to fine-grained event boundaries than coarse-grained event boundaries (Zacks, 2004). Does this hold for naturalistic action? To answer this question, Experiment 1 manipulated the grain at which viewers segmented activity.

Finally, these studies investigated how conceptual information interacts with movement information to determine event segmentation. Previous studies have found that providing comprehenders with a conceptual frame for an activity before reading about it can profoundly affect comprehension (Bransford & Johnson, 1972; George, Kutas, Martinez, & Sereno, 1999; Maguire, Frith, & Morris, 1999). Conceptual framing can affect processes that are generally thought to be fast and early in comprehension, including the resolution of lexical ambiguity (Wiley & Rayner, 2000). Conceptual framing has also been shown to affect event comprehension and segmentation (Massad, Michael, & Newtson, 1979). Does one's conceptual frame affect how one processes movement information to extract meaning? One way of describing conceptual representations of events is in terms of *schemata*, which are structured knowledge representations of types of things and events that one has encountered in the past (Rumelhart, 1980). Schemata represent typical feature values for a type of entity and relations amongst those features. For example, a schema for "folding laundry" might include information about the sorts of objects that are typically involved (e.g., clothes, baskets) and the order in which steps are typically performed. If a viewer familiar with laundry-folding views a movie that shows a pile of clothes, a basket, and a person performing steps such reaching into the basket, the viewer might well activate their schema for folding laundry, as well as schemata corresponding to the different object types present, and a person schema.

Schema activation can have two distinct effects on ongoing perception. First, active schemata provide new information. Activating a schema for folding laundry provides information about what objects are likely to be present, what steps are likely to be performed, and in which order. If these additional information sources play a role in event segmentation, then their presence might weaken effects of movement variables on segmentation. Second, active schemata can change how information is processed, by biasing processing or modulating attention. For example, imagine the viewer sees an ambiguous, partly occluded motion pattern that is consistent with the actor folding a towel in half or wringing it out. In the context of an active laundry-folding schema that sensory signal might receive an interpretation consistent with folding, whereas in the context of an active kitchen-cleaning schema the same signal might receive an interpretation consistent with wringing out. Another example: activating a laundry-folding schema might increase attention to the distance between the hands, whereas activating a hair-combing schema might increase attention to the distance between the dominant hand and the head. Thus, activating a schema for an activity may affect not just the weight given to movement information, but also how that information is processed. This would be expected to affect *which* movement features are correlated with segmentation.

Thus, conceptual framing could have two distinct effects on the relations between movement and event segmentation: Weakening the overall strength of relations between movement and segmentation, and changing which features of movement are correlated with segmentation. These two possibilities are not mutually exclusive. Experiment 2 investigated these two potential effects of conceptual framing by manipulating how much information viewers had about variables other than movement before they segmented an activity and while they were segmenting.

In short, the experiments reported here aimed to answer four questions about the role of movement information in event segmentation. First, does movement quantitatively predict segmentation of naturalistic activity? Second, does removing some of the visual information provided by naturalistic videos increase viewers' dependence on movement features for segmentation? Third, do the relations between movement and segmentation depend on the grain at which the viewer segments? Finally, do the relations between movement and segmentation depend on conceptual framing?

2. Experiment 1

In Experiment 1, participants segmented movies of everyday activities performed by a single actor using a set of objects on a tabletop. The actor's movements were recorded with a magnetic tracking system, allowing us to analyze the relations between movement variables and event segmentation in naturalistic activity. Viewers segmented the activity at both a fine and coarse grain, allowing us to ask whether movement was more strongly related to fine-grained than coarse-grained segmentation.

2.1. Method

2.1.1. Participants

Twenty-six students at Washington University (ages 18–22, 20 female) participated in partial fulfillment of a course requirement. An additional five participants failed to complete the experiment due to computer problems (two), illness (one), or failure to follow the instructions (two).

2.1.2. Stimuli

Participants watched three movies of a college-aged man performing everyday tabletop activities—folding laundry (498 s), building a house from Duplos (371 s; Lego Group, www.lego.com), and assembling a video game system (240 s). For training, the initial 180 s of a movie showing the man assembling a cardboard shelving unit was used. The movies were filmed from a fixed head-height perspective using a digital camera and reduced to 320×240 pixel resolution for display. All movies began and ended with several seconds of the actor sitting still. Examples of the stimuli are shown in Fig. 1. The complete videos are available at http://dcl.wustl.edu/DCL/Stimuli.html.

During filming, the actor was outfitted with three magnetic sensors to record the position of his hands and head. The hand sensors were attached to the back of the hands with medical tape; the head sensor was attached to the rear top of the head using tape and a woolen cap (see Fig. 1). The motion tracking apparatus (Flock of Birds, Ascension Technologies, Burlington VT) was controlled by a PC and synchronized with the video recording after data acquisition. Positions were recorded at 29.27 Hz. (During recording of the videogame event, the system lost signal from 159.9 to 170.5 s in the movie. These frames were excluded from analysis.)

2.1.3. Segmentation task

Participants segmented each movie to identify boundaries between events. They were told that they would be watching movies of everyday activities and that they should press a button on a button box whenever, in their judgment, one natural and meaningful unit of activity ended and another began. They were told that the actor would be wearing sensors to track the positions of his head and hands, and that they could ignore the sensors. Each

Fig. 1. Still frames taken from the laundry (top), Duplos (middle), and videogame (bottom) movies used in Experiments 1 and 2.

participant segmented all three movies twice, once to mark *coarse* event boundaries and once to mark *fine* boundaries. For coarse segmentation they were asked to identify the largest units of activity that were meaningful to them. For fine segmentation they were asked to identify the smallest units of activity.

Movies were presented on a Macintosh computer (www.apple.com) with a 19 in monitor, using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Responses were recorded using the PsyScope button box.

2.1.4. Procedure

Each participant was given either fine-grained or coarse-grained instructions for the segmentation task and then trained using the 180 s practice movie. The experimenter offered to answer any questions. The participant then segmented the three stimulus movies. This procedure was repeated for the other segmentation grain. Order of segmentation grain and movie order was counterbalanced across participants.

2.2. Results

2.2.1. Movement analysis

We analyzed the motion tracking recordings to provide a record of the actor's movement over time. First, the transient data collection errors were corrected by visual inspection. Next, a set of 15 variables describing the actor's movement were calculated from the position information:

- the speed of each hand and the head,
- the acceleration of each hand and the head,
- the pairwise distance between each of the three tracked points (left hand, right hand, head),
- the pairwise relative speed, and
- the pairwise relative acceleration.

For example, if the actor were resting his left hand on his head and then began to move it toward the table, the pairwise left-hand-to-head distance would increase, the pairwise left-hand-to-head speed would become positive (indicating that distance was increasing over time), and the pairwise left-hand-to-head acceleration would become positive (indicating that the rate of increase of distance was increasing over time). As the hand reached the table the pairwise distance would change more slowly, the pairwise speed would approach zero, and the pairwise acceleration would pass through zero, become negative, and then reach zero again. The movement variables were then resampled to a 1 Hz rate for comparison with the behavioral data, using kernel estimation with a 1-s bandwidth.

2.2.2. Relationship between movement and segmentation

To compare participants' segmentation to the movement variables, we first binned each participant's segmentation data to 1 s intervals. For each interval in each movie we then counted the number of participants who identified a fine event boundary and the number who identified a coarse boundary, producing two time series. As can be seen in the example in Fig. 2, event boundaries were clustered such that some intervals were marked as boundaries by a high proportion of participants and others were marked as boundaries by few participants. These measures of frequency of segmentation over time were then compared to the movement variables.

One simple and intuitive measure of the strength of relationship between a movement variable and segmentation frequency is the correlation coefficient, *r*. For example, a positive correlation between head speed and segmentation indicates that participants tended to segment when the head was moving rapidly. However, the simple correlation presumes that the movement variables and segmentation are perfectly in phase. This assumption is likely not

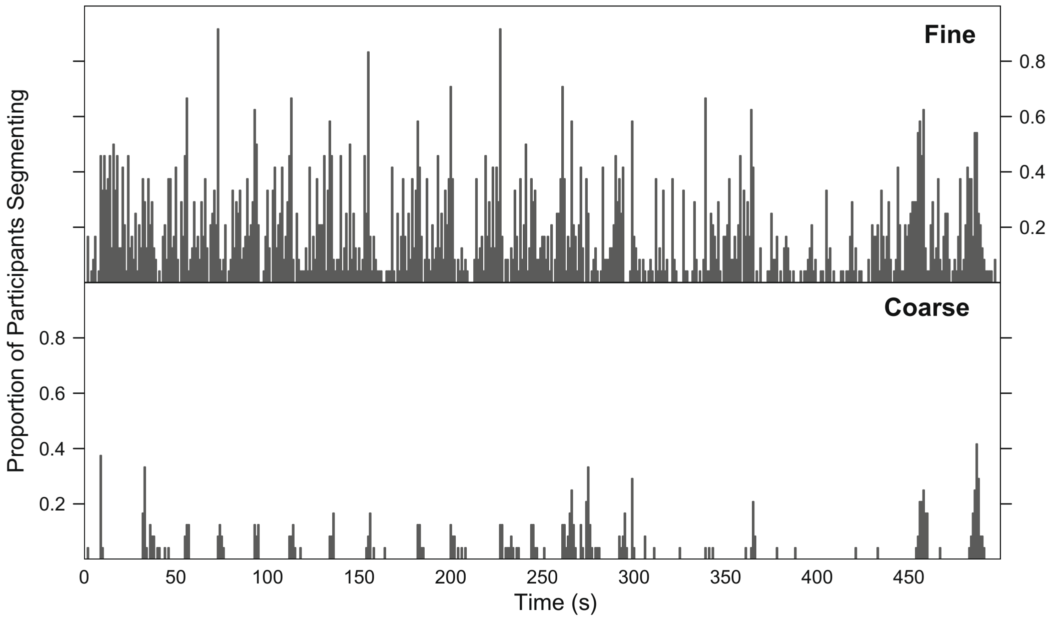

Fig. 2. Proportion of participants who identified a coarse or fine event boundary during each 1-s interval of the laundry folding movie.

Table 1
Correlations between movement variables and event segmentation after shifting the movement variables by up to 5 s to maximize the absolute value of the correlations. Values are means across movies (SDs in parentheses).

	Experiment 1 (fine and coarse)		Experiment 2a (fine)		Experiment 2b (coarse)	
	Largest correlation	Optimal lag	Largest correlation	Optimal lag	Largest correlation	Optimal lag
Speed						
Right hand	0.19 (0.08)	−0.33 (0.58)	0.44 (0.10)	0 (0)	0.28 (0.10)	0.33 (0.52)
Left hand	0.33 (0.08)	−0.33 (0.58)	0.66 (0.05)	0 (0)	0.38 (0.05)	0 (0)
Head	0.31 (0.1)	−0.33 (0.58)	0.56 (0.08)	0 (0)	0.46 (0.07)	0 (0)
Acceleration						
Right hand	0.16 (0.06)	0 (1)	0.34 (0.07)	0.33 (0.52)	0.25 (0.12)	−0.33 (2.58)
Left hand	0.24 (0.04)	−0.33 (0.58)	0.49 (0.12)	0 (0)	0.25 (0.10)	−0.33 (1.97)
Head	0.3 (0.07)	−0.33 (0.58)	0.52 (0.11)	0 (0)	0.42 (0.08)	0 (0)
Distance						
Right hand–left hand	0.26 (0.06)	−0.33 (1.53)	0.44 (0.09)	0 (0)	0.26 (0.05)	0 (0)
Right hand–head	0.18 (0.03)	0.33 (1.15)	0.18 (0.20)	0.5 (2.35)	0.04 (0.23)	−0.83 (2.79)
Left hand–head	0.26 (0.09)	−0.33 (0.58)	0.36 (0.10)	−0.33 (0.52)	0.22 (0.12)	0.17 (0.75)
Relative speed						
Right Hand–Left Hand	−0.08 (0.15)	−0.33 (1.53)	0.13 (0.16)	0.67 (82)	0.11 (0.13)	1.33 (1.97)
Right hand–head	−0.14 (0.04)	−1.33 (0.58)	−0.01 (0.14)	0.33 (2.16)	0.04 (0.18)	0.17 (3.66)
Left hand–head	−0.13 (0.02)	−2.33 (1.53)	0.06 (0.17)	0.17 (0.98)	0.07 (0.13)	−0.17 (0.75)
Relative acceleration						
Right hand–left hand	0.02 (0.09)	1.67 (2.52)	−0.17 (0.09)	0.17 (0.41)	−0.08 (0.11)	−0.67 (1.97)
Right hand–head	−0.03 (0.07)	−1.67 (1.15)	0.05 (0.09)	1.67 (2.42)	−0.03 (0.13)	0.50 (2.66)
Left hand–head	−0.04 (0.11)	1.67 (1.53)	−0.07 (0.14)	−0.17 (1.33)	−0.13 (0.05)	0.17 (2.79)

Note: Positive lag indicates that segmentation frequency correlated most strongly with subsequent values of a movement feature; negative lag indicates segmentation frequency correlated most strongly with previous values of a movement feature.

warranted. For example, suppose a viewer tried deliberately to segment when the actor's hands were maximally outstretched (i.e., at local maxima in the right hand–left hand distance). The viewer could not know precisely when these maxima have occurred until after they are over. Thus, the effects of hand distance on segmentation would have a temporal lag. To account for these phase relationships we fitted cross-correlation sequences between each movement variable and segmentation frequency for each movie (Zacks, 2004). Coarse and fine segmentation were combined by dividing each frequency time series by its standard deviation (to compensate for greater numbers of boundaries in the fine segmentation condition) and summing the two time series. We then calculated the cross-correlation between the summed segmentation frequency series and each movement variable, using lags from −5 to 5 1-s bins, and noted the lag with the largest correlation. The lag and correlation were stored, and the movement variable was shifted in time using the noted lag to maximize the correlation between the movement variable and the segmentation frequency series. The means across movies of the highest correlations and optimal lags are given in Table 1. As can be seen in the table, the speed and acceleration of the left hand and head were consistently positively correlated with segmentation, as were the distances between the two hands and between each hand and the head. For all of these features the mean lags were small, between −.33 and .33. Larger lags tended to be associated with smaller optimal correlations; this makes sense because when the cross-correlation sequence has no strong maximum the estimate of the optimal lag will be variable. The most frequently occurring lag was zero (16 of 45), followed by −1 (14 of 45) and 1 (5 or 45), indicating that effects of changes in movement variables on segmentation were generally seen during the same 1-s interval as the change or during the following interval.[1]

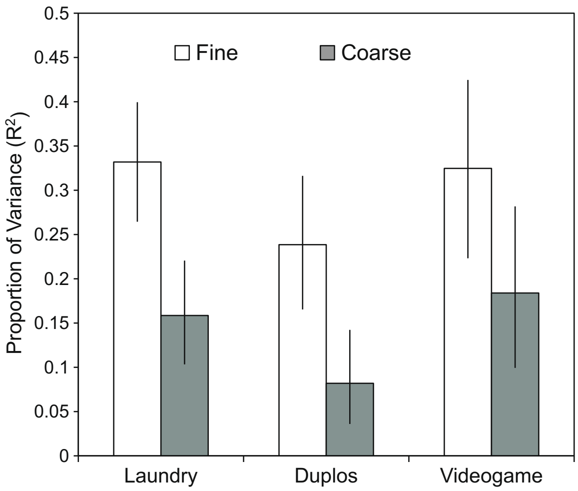

Fig. 3. Movement variables accounted for substantial variance in event segmentation in Experiment 1, particularly for fine-grained segmentation. (Error bars are 95% confidence intervals.)

With these optimally shifted movement variables in hand, we performed multiple linear regression analyses

[1] Positive lags indicate that the effect of a change in a movement variable on segmentation is seen before the change in the movement feature itself. This at first appears paradoxical, but can occur due to autocorrelation in the movement variables.

Table 2
Movement variables that were significantly correlated with event segmentation for each combination of movie and segmentation grain ($p < .05$ corrected for multiple comparisons across movement features).

	Coarse			Fine		
	Videogame	Duplos	Laundry	Videogame	Duplos	Laundry
Speed						
Right hand				+	+	+
Left hand			+	+	+	+
Head			+	+	+	+
Acceleration						
Right hand			+	+	+	+
Left hand				+	+	+
Head	+		+	+	+	+
Distance						
Right hand–left hand		+		+	+	+
Right hand–head						+
Left hand–Head			+	+	+	+
Relative speed						
Right hand–left hand				−		
Right hand–head					−	
Left hand–head						−
Relative acceleration						
Right hand–left hand						
Right hand–head						
Left hand–head						−

+: Relationship between movement variable and segmentation was positive.
−: Relationship between movement variable and segmentation was negative.

to answer three questions: First, how strong was the overall relationship between movement variables and event segmentation for these movies? Second, did the strength of this relationship change with segmentation grain? Third, which particular movement variables were predictive of segmentation? For each movie, we fit two linear regressions, one predicting the proportion of participants who identified a coarse boundary during each 1-s interval, and another predicting the proportion of participants who identified a fine boundary. The predictors for both models were the 15 optimally shifted movement variables. We took the total variance accounted for in the regression (R^2) as a measure of the total strength of the relation between movement and segmentation, and examined the individual regression coefficients to characterize the sign and strength of the relation for individual movement features.

As can be seen in Fig. 3, movement variables were significant predictors of segmentation, accounting for 8% to 33% of the variance in segmentation. For all three movies the relationship between movement variables and segmentation was substantially stronger for fine segmentation than for coarse segmentation.[2] Table 2 indicates which features were significantly correlated with event segmentation for each combination of movie and segmentation grain. The most consistent predictors of segmentation were the speed and acceleration of body parts and the distance between the left hand and other body parts. These relationships were always positive, indicating that participants tended to segment when the hands and head were accelerating or moving quickly, and when the left hand was far from the right hand or head. (It is worth noting that the actor was left-handed.)

2.2.3. Event unit lengths

The lengths of the units participants identified are given in Table 3. As can be seen in the table, participants were able to modulate their segmentation grain as instructed, identifying larger units for coarse-grained segmentation and smaller units for fine-grained segmentation. The table also indicates that the units identified in the videogame movie were generally shorter than those in the other movies, particularly for coarse-grained segmentation. These patterns led to significant main effects of segmentation grain [$F(1, 25) = 76.2$, $p < .001$] and movie [$F(2, 50) = 13.1$, $p < .001$], as well as a significant interaction [$F(2, 50) = 11.0$, $p < .001$].

2.2.4. Hierarchical organization

The degree to which viewers grouped fine events hierarchically into coarse events was assessed using the *enclosure* measure proposed by Hard, Recchia, & Tversky (submitted for publication), and using the *alignment* measure proposed by Zacks, Tversky, & Iyer (2001). Enclosure is calculated by comparing each coarse event boundary to the nearest fine boundary identified by the same participant. If viewers spontaneously group fine-grained events into coarse-grained events, the majority of these nearest fine boundaries should fall before the coarse boundary to which they are closest. Therefore, enclosure is calculated as the proportion of coarse boundaries that fall after their

[2] We repeated the analyses using optimal lags computed separately for fine and coarse segmentation. The results were similar, with movement features still predicting fine segmentation more strongly than coarse segmentation.

Table 3
Coarse and fine unit lengths in seconds as a function of movie and segmentation grain in Experiment 1. Values are means across participants (SDs in parentheses).

Experiment 1	Coarse	Fine
Videogame	33.56 (14.33)	8.52 (6)
Duplos	62.60 (41.94)	10.51 (8.84)
Laundry	60.89 (41.54)	11.24 (8.55)

nearest fine boundary. Enclosure scores greater than 0.5 indicate hierarchical organization. The mean enclosure score was 0.59 (SD across participants = .13), which differed significantly from 0.5, $t(25) = 23.7$, $p < .001$. There were no significant differences in enclosure across the three movies, $F(2, 50) = .04$, $p = .96$.

Alignment measures the degree to which coarse event boundaries correspond to a subset of fine boundaries. It is calculated by measuring the distance from each coarse event boundary to its nearest fine event boundary, and comparing those distances to that which would be expected if there were no relationship between the locations of coarse and fine boundaries. Observed distances had a mean of 1.67 s (SD across participants = 1.39), whereas the null expectation was 4.82 s (SD = 3.72 s), $t(25) = -5.49$, $p < .001$. The difference was smaller for the videogame movie (M = 2.15 s, SD = 2.55 s) than for the Duplos movie or the laundry movie (M = 3.75 s, SD = 3.41 s, and M = 3.56 s, SD = 4.35 s, respectively), and this effect of movie was statistically significant, $F(2, 50) = 3.48$, $p = .04$. This likely reflects a scaling effect; the videogame movie produced finer-grained segmentation overall, leading to smaller actual and observed distances.

2.3. Discussion

This experiment provided a clear answer to our first question: Movement variables were robustly correlated with event segmentation when viewing naturalistic everyday activities. Viewers were more sensitive to the movements of individual body parts and the distance between them than to the relative speed and acceleration of the body parts with respect to each other. This may reflect that body part movements are coded with respect to a common trunk-centered reference frame (Johansson, 1973) rather than in terms of effector-to-effector relations. Viewers were more sensitive to movements of the left hand than of the right hand. This may reflect the handedness of the actor—being left-handed, it is likely that he provided more informative cues with the left than the right hand. Another possibility, which we think less likely, is that the left hand was processed more thoroughly because it was generally closer to the camera in the shots we selected. In future studies it would be of interest to systematically vary the handedness of actors and the orientation of the camera.

The experiment also clearly answered our second question: Movement variables were better predictors of fine-grained segmentation than coarse-grained segmentation. This replicates previous findings using simple animations (Zacks, 2004), and is consistent with the view that fine events are particularly focused on individual actions on objects.

3. Experiments 2a and 2b

Experiments 2a and 2b investigated the interaction between perceptual and conceptual information in event segmentation. Specifically, they were designed to answer two questions: First, does removing other cues to event segmentation strengthen the relations between movement variables and segmentation? Second, does one's prior conceptual representation of an activity affect the ongoing processing of movement information?

To test both of these possibilities, we manipulated the degree to which viewers had information about the activity being performed, the objects being interacted with, and the actor's gaze and expression. In the *video* condition, participants segmented movies while watching live-action movies as in Experiment 1. For the *animation-informed* condition, we created animations of the actor's hands and head from the motion tracking data. Participants segmented these animations, but before doing so viewed a 40-s preview of the live-action video. Finally, participants in the *animation-uninformed* condition viewed the same animations but without the live-action preview. In all three conditions, observers had access to the movements of the head and hands. The conditions were designed to vary in the additional information present. Compared to the video condition, the animation condition was designed to deprive participants of two potential bases for segmentation: conceptual knowledge about the activity being performed (schemata) and visual features other than the motion of the head and hands. In the video condition observers should have been able to easily recognize the activity being performed and thus to activate relevant event schemata, cued by the objects present and the actor's interaction with those objects. They also had ongoing access to a number of visual features beyond the motion of the head and hands: Videos provide ongoing information about objects' identities and locations, the actor's contact with those objects, and the actor's facial expression and gaze. Videos also provide much information about movement features not captured by the animations—for example, the angular motion of the head, the relative movements of the fingers, and the movements of the elbows. We predicted that participants in the animation conditions would show stronger relations between movement and segmentation than those in the video condition, because they would have access to fewer additional features that might affect their segmentation.

We also hypothesized that if one's prior conceptualization of an activity affects how movement features are processed, then the two animation groups should differ in the relation between their segmentation and movement features. The animation-informed condition was specifically designed to provide observers an opportunity to activate relevant schemata (during the 40-s preview) while equating the visual features present during segmentation with the animation-uninformed condition. Because participants segmented identical stimuli in the two animation conditions, any differences in their segmentation patterns would likely be due to conceptual representations formed by the animation-informed group during the preview stage. Prior conceptualization, if present, should affect both the magnitude and nature of the relations between movement features and segmentation. First, it should change how strongly movement features predict segmentation. We hypothesized that if activating a schema for an activity provided additional conceptual features that fed into event segmentation, this would render movement features less strongly related to segmentation. Second, prior conceptualization should change which movement features predict segmentation, rendering features that are more schema-relevant more predictive and features that are less schema-relevant less predictive.

The effect of conceptual representations on event perception might differ for different grains of segmentation. One possibility is that fine-grained segmentation is more perceptually driven, whereas the grouping of fine-grained units into larger structures is more conceptually driven. If so, one would expect the animation-informed and animation-uninformed groups to differ more for coarse-grained segmentation than for fine-grained segmentation. On the other hand, fine-grained segmentation appears to be related more strongly to movement features than does coarse-grained segmentation (Experiment 1; Zacks, 2004). Therefore, if conceptual structure modulates the relationship between movement and segmentation, then differences between the animation-informed and animation-uninformed groups might be more apparent in fine-grained segmentation. To test these possibilities, Experiment 2a measured fine-grained segmentation whereas Experiment 2b measured coarse-grained segmentation.

A secondary goal of these experiments was to replicate the primary finding of Experiment 1—a robust relation between movement variables and event segmentation in naturalistic action—using a larger stimulus set. To that end we tested participants on six everyday activities. With this larger stimulus set, task fatigue and boredom were a concern, so segmentation grain was manipulated between participants to reduce the session length. Fine-grained segmentation was tested first, in Experiment 2a, and coarse-grained segmentation was tested second, in Experiment 2b.

3.1. Method

3.1.1. Design

Experiments 2a and 2b were run sequentially; however they will be considered together as one study for most analyses. Viewed this way, there were two independent variables, both manipulated between participants. *Grain* of segmentation was fine in Experiment 2a and coarse in Experiment 2b. *Stimulus condition* was manipulated between participants within each experiment by randomly assigning each participant to either the video, animation-informed, or animation-uninformed stimuli and instructions.

3.1.2. Participants

Fifty-four students at Washington University participated in each experiment in partial fulfillment of a course requirement (Experiment 2a: ages 18–22, 40 female; Experiment 2b: ages 18–22, 39 female). An additional two participants in Experiment 2a declined to complete the experiment and were excused; one additional participant in Experiment 2b was unable to complete the protocol due to experimenter error.

3.1.3. Stimuli

The three stimulus activities from Experiment 1 were used again in Experiment 2, but the durations of the intervals before and after the actor appeared were changed slightly in editing, resulting in slightly different movie durations: 501 s for folding laundry, 380 s for building a house from Duplos, and 245 s for assembling a videogame system. In addition, three new activities were used: paying bills (388 s), improvising an abstract structure with Duplos (365 s), and making a peanut butter and jelly sandwich (332 s).

For each of the activities, an animation was constructed by rendering the left and right hands as reflective green and red balls, respectively, and the head as a reflective blue ball (see Fig. 4). The left and right hand balls were connected to the head ball with thin gray rods whose length varied as the distance between the hands and head changed. To maximize accurate depth perception, the scene was rendered with a ground plane corresponding to the tabletop height, onto which the balls cast shadows. The animations were created with raster3d (http://www.bmsc.washington.edu/raster3d). Both the live-action movies and animations were presented at 720×480 resolution. Animations were rendered at 10 fps, which was sufficient to produce smooth-appearing motion; live-action movies were displayed at their native 29.97 fps.

In this version of the videogame stimulus, the motion tracking acquisition failure (see Experiment 1 Method) be-

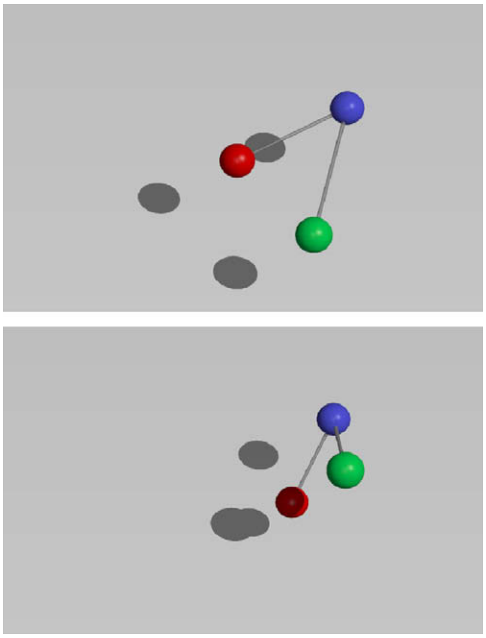

Fig. 4. Two still frames from the laundry animation used in Experiment 2.

gan at 163.9 s and ended at 174.6 s. During this interval the animation showed a large red X instead of the three balls. Data from this interval and the following 70.1 s were excluded from all analyses.

3.1.4. Segmentation task

Participants performed the segmentation task as in Experiment 1. Rather than segmenting twice to mark fine and coarse units, each participant segmented once. In Experiment 2a they were given fine-grained segmentation instructions; in Experiment 2b they were given coarse-grained instructions.

3.1.5. Procedure

Each participant was assigned to one of the video, animation-uninformed, or animation-informed stimulus conditions. Participants in the video condition were trained as were participants in Experiment 1, using the same practice activity. They were told that they would be viewing a movie of an actor engaged in an everyday activity, and that the objects attached to his hands and his head are there to record his movements. Participants in the animation-uninformed condition were told that they would be viewing a video of the movements of three colored balls connected by two rods, and that they should interpret the movements as being meaningful and goal-directed. Participants in the animation-informed condition were told that they would be viewing animations generated from the movements of an actor's head and hands. For the practice movie, the animation was superimposed on the corresponding live-action movie, in the upper left corner, to illustrate that the rendered points corresponded to the actor's head and hands.

Participants then went on to segment the six target movies, with order counterbalanced across participants. For participants in the animation-informed condition, each animation to be segmented was preceded by 40 s of the corresponding live-action movie. This was done to maximize the degree to which viewers would be able to form a vivid image of the actions performed from the movements of the balls in the animation.

3.2. Results

For all analyses, data from the first 40 s of each movie were excluded, because participants in the animation-informed condition had previously viewed the live-action video corresponding to those 40 s.

3.2.1. Relationship between movement and segmentation

Movement information was analyzed using the same procedures as for Experiment 1. The movement information was captured, filtered, and resampled to a 1-s sampling rate. The segmentation data were binned to 1 s intervals and the segmentation counts for each interval were averaged (separately for each group) to estimate the frequency of segmentation over time.

For each movement feature we calculated the cross-correlation between that feature and the combined segmentation data for the three groups; this was performed separately for each feature and each movie, and separately for Experiments 2a and 2b. As can be seen in Table 1, correlations were overall higher than those in Experiment 1. This probably reflects the fact that with a larger number of participants estimates of segmentation are more reliable. The pattern of correlation across features was quite similar to that in Experiment 1: Speed and acceleration were again strong predictors, particularly those of the left hand, followed by the distances between the left hand and the head and between the left and right hands. As in Experiment 1, the most frequently occurring best-fitting lag was 0 (59 of 90 in Experiment 2a, 50 of 90 in Experiment 2b), followed by 1 (14 of 90 in Experiment 2a, 10 of 90 in Experiment 2b) and -1 (8 of 90 in Experiment 2a, 6 of 90 in Experiment 2b). Larger lags occurred mostly with small (i.e., unreliable) correlations.

As for Experiment 1, we performed multiple linear regression analyses with segmentation frequency as the dependent measure and the 15 movement variables as the predictors. Regressions were performed separately for each movie and each group. As can be seen in Fig. 5, movement variables were again strong predictors of segmentation frequency. Replicating Experiment 1, movement predicted fine segmentation (Experiment 2a) more

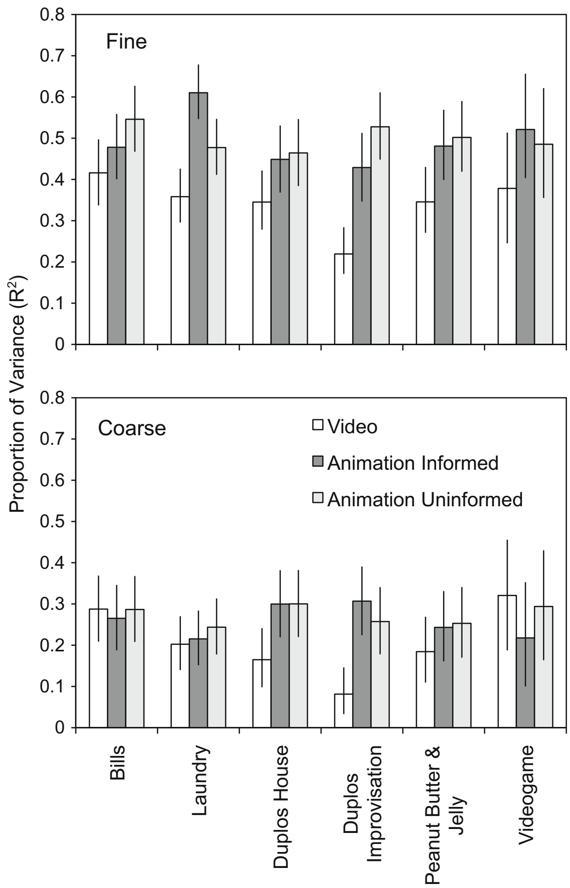

Fig. 5. Movement variables accounted for substantial variance in event segmentation in Experiments 2a and 2b. This was stronger for fine segmentation (Experiment 2a, top) compared to coarse segmentation (Experiment 2b, bottom), and stronger for the animation conditions than for the video condition. (Error bars are 95% confidence intervals.)

strongly than coarse segmentation (Experiment 2b), leading to higher R^2 values. To test whether the difference in R^2 between coarse and fine segmentation was statistically robust, and to assess whether R^2 differed across the stimulus conditions, we conducted an analysis of variance (ANOVA) with the R^2 value from each regression as the dependent measure, grain and stimulus condition as the independent measures, and movie as a blocking variable. The difference between fine and coarse grains was statistically significant, $F(1, 25) = 122.6$, $p < .001$. Movement also predicted segmentation more strongly in the two animation conditions than in the video conditions, $F(2, 25) = 15.4$, $p < .001$. However, the effect of stimulus condition was qualified by a marginally significant grain-by-stimulus condition interaction, $F(2, 25) = 3.10$, $p = .06$. To follow this up we conducted ANOVAs separately for Experiments 2a and 2b. For Experiment 2a there was a significant effect of stimulus condition, $F(2, 10) = 16.3$, $p < .001$, whereas for Experiment 2b this effect was not significant, $F(2, 10) = 2.0$, $p = .19$. To test the specific hypothesis that viewers' conceptual frames affected the strength of the relationship between movement and segmentation, we compared the R^2 statistics for the two animation groups in each experiment. In neither case was this difference statistically significant [Experiment 2a: $t(5) = 0.74$, $p = .87$; Experiment 2b: $t(5) = 0.87$, $p = .43$].

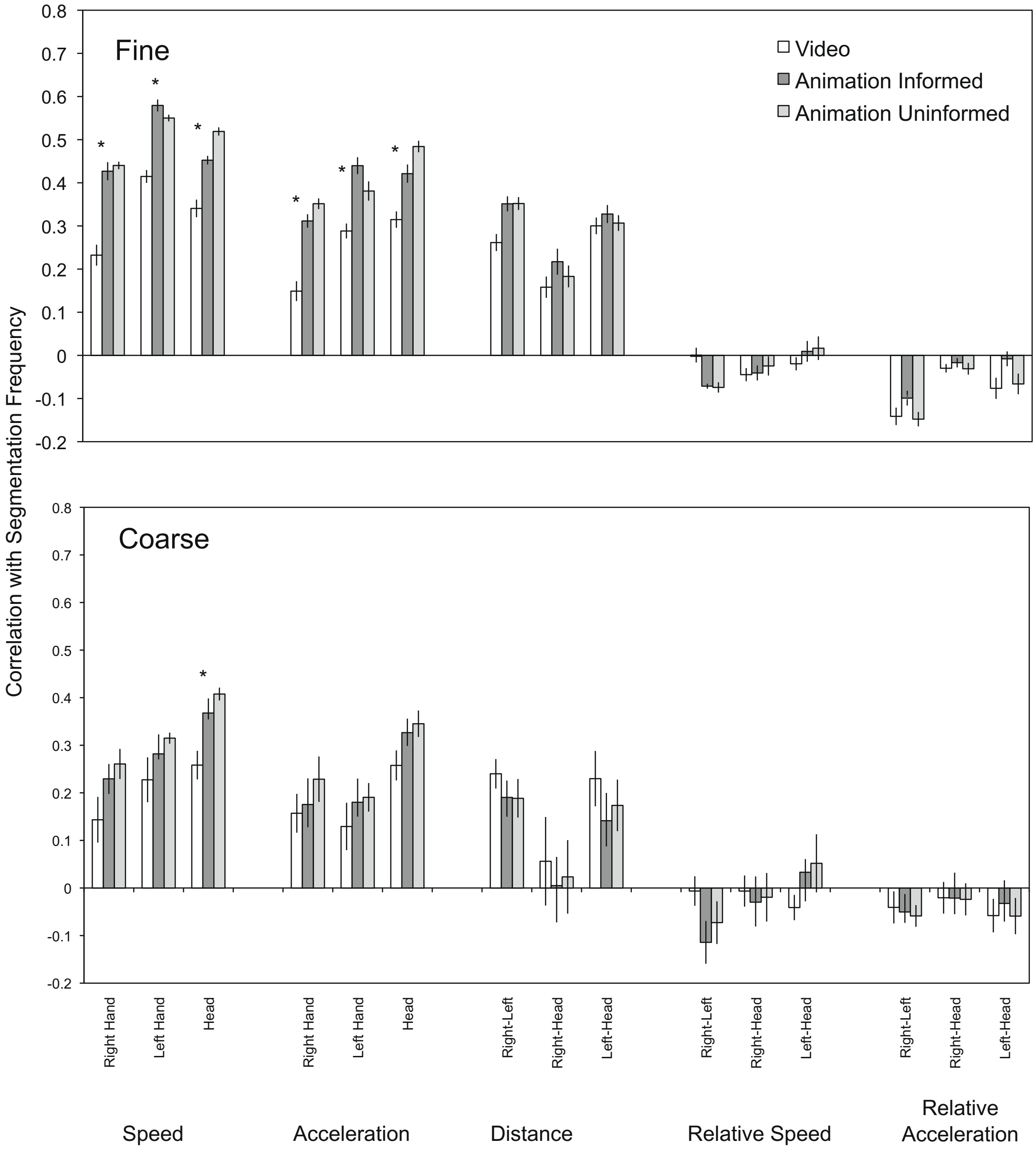

Fig. 6. Speed, acceleration and distance were positively correlated with segmentation in Experiment 2, particularly for the animation conditions. (Error bars are standard errors of the mean. Asterisks mark features for which there was a significant group difference, corrected for multiple comparisons across features.)

To assess the strength of the relationship between individual movement features and segmentation, we computed correlations between each movement feature separately for each movie for each group in each of the two experiments. The results are plotted in Fig. 6. Similar to Experiment 1, the speed and acceleration of body parts, and the distances between them, were the best predictors of segmentation. Correlations for speed and acceleration again were positive, indicating that viewers segmented when body parts were moving quickly and speeding up. Correlations for distances generally again were positive, indicating that viewers segmented when body parts were far apart. Similar groups of features were related to segmentation across experimental conditions. To test these statistically, we converted the correlations to normally distributed variables with Fisher's z transformation and subjected them to one group t-tests for each movement variable, with a Bonferroni correction across the 15 features for each of the two experiments. The speed and acceleration of the hands and head were significantly correlated with segmentation [Experiment 2a: smallest $t(35) = 8.36$, corrected $p < .001$; Experiment 2b: smallest $t(35) = 9.71$, corrected $p < .001$]. The distance between the head and each hand, and the distance between the hands, were all significantly correlated with segmentation, [Experiment 2a: smallest $t(35) = 5.07$, corrected $p < .001$; Experiment 2b: smallest $t(35) = 3.51$, corrected $p = .001$]. Finally, the relative speed of the two hands, and their relative acceleration, both were negatively correlated with segmentation. The correlation for relative speed was not significant for Experiment 2a [$t35) = -2.06$, corrected $p = .83$] but was significant for Experiment 2b [$t(35) = -3.55$, corrected $p = 0.001$. The correlation for relative acceleration was significant for both experiments [Experiment 2a: $t(35) = -5.18$, corrected $p = .001$; Experiment 2b: $t(35) = -5.45$, corrected $p < .001$].

To test whether stimulus condition affected how strongly individual movement variables correlated with event segmentation, we conducted a set of univariate ANOVAs with stimulus condition as the independent measure and movie as a blocking variable. This analysis was conducted separately for the fine (Experiment 2a) and coarse (Experiment 2b) segmentation groups. For each ANOVA the dependent measure was the Fisher-transformed correlation between a movement variable and segmentation frequency. The results are indicated with asterisks in Fig. 6. As can be seen in the figure, for fine segmentation stimulus condition significantly affected the strength of the correlation between segmentation frequency and six variables: the speed and acceleration of the left and right hands and the head. In all cases, follow-up t-tests indicated that the video group had a significantly smaller correlation than one or both of the animation groups. There were no group differences in the correlations of segmentation frequency with distances between body parts, or with relative speed or acceleration of body parts. For coarse segmentation the only effect of stimulus condition on the correlations was for the speed of the head. Follow-up t-tests indicated that the two animation groups had significantly higher correlations than the video group.

In sum, movement variables were again strongly correlated with event segmentation. Movement was more strongly related to segmentation for those who segmented at a fine grain, particularly when they watched animations rather than live-action video. Participants were more likely to identify event boundaries (a) when the head and hands were moving quickly, (b) when the speed of the head and hands was increasing, (c) when the head and hands were far apart, (d) when the hands were moving toward each other, and (e) when the hands were accelerating toward each other. For neither fine nor coarse segmentation was there evidence that the two animation groups differed from each other in the strength of the relationship between movement and segmentation or in which movement features predicted segmentation. Therefore, there was no evidence for an effect of conceptual frame on segmentation.

3.2.2. Event unit lengths

As can be seen in Table 4, unit lengths in Experiments 2a were comparable to those from the fine segmentation condition in Experiment 1, and unit lengths from Experiment 2b were comparable to the coarse condition in Experiment 1. A mixed ANOVA with segmentation grain and stimulus condition as between-participants variables and

Table 4
Coarse and fine unit lengths in seconds as a function of movie and segmentation grain in Experiments 2a and 2b. Values are means across participants (SDs in parentheses).

	Video	Animation informed	Animation uninformed
Fine segmentation (Experiment 2a)			
Bills	12.84 (8.08)	9.05 (5.72)	11.56 (10.31)
Laundry	14.34 (9.04)	7.10 (6.39)	9.66 (9.09)
Duplos house	9.77 (8.66)	5.20 (5.32)	6.88 (8.08)
Duplos improvisation	11.41 (9.96)	6.56 (4.92)	7.56 (5.96)
Peanut butter and jelly	9.29 (5.48)	8.34 (6.06)	9.19 (8.25)
Videogame	8.23 (4.45)	7.05 (3.46)	8.84 (6.49)
Coarse segmentation (Experiment 2b)			
Bills	45.97 (25.48)	24.43 (15.02)	23.85 (10.78)
Laundry	70.64 (65.33)	26.20 (20.01)	25.85 (24.35)
Duplos house	73.95 (47.46)	22.85 (14.66)	22.89 (18.72)
Duplos improvisation	59.21 (25.80)	21.76 (15.31)	25.03 (24.34)
Peanut butter and jelly	46.18 (22.38)	20.89 (10.05)	29.52 (26.33)
Videogame	31.46 (12.61)	18.99 (14.13)	16.98 (7.31)

movie as a repeated measure indicated that coarse units were significantly larger than fine units, as expected, $F(1, 102) = 70.51$, $p < .001$. The video group identified larger units than the two animation groups, particularly for coarse segmentation, leading to a significant main effect of stimulus condition [$F(2, 102) = 15.19$, $p < .001$] and a significant grain-by-stimulus condition interaction [$F(2, 102) = 10.41$, $p < .001$]. As in Experiment 1, unit lengths varied across the movies: There was a significant main effect of movie and all the interactions involving movie were significant, indicating that the grain and stimulus condition effects varied across the movies (smallest $F = 3.44$, $p < .001$). To better characterize the grain-by-stimulus condition interaction, we evaluated the main effect of stimulus condition in separate follow-up ANOVAs for each of the two experiments. For fine segmentation (Experiment 2a), there was no significant effect of stimulus condition, $F(2, 51) = 1.51$, $p = .23$. For coarse segmentation, the effect of stimulus condition was significant, $F(2, 51) = 13.93$, $p < .001$. In sum, participants identified somewhat larger event units from videos than from animations. This was particularly true when they identified coarse-grained events.

3.3. Discussion

3.3.1. Experiments 2a and 2b provided a robust replication of the primary result of Experiment 1

Movement variables were strongly correlated with segmentation of naturalistic everyday activities. This experiment also provided a clear answer to our question as to whether movement would be more strongly related to segmentation when other cues to segmentation were removed: The two groups who viewed relatively impoverished animations showed stronger correlations between movement variables and segmentation than did the group who viewed the live-action videos. We interpret this as indicating that the segmentation of the video groups depended on movement features as in the animation conditions, and also on the additional features that video provides. This may include information about objects' identities and locations, information about the actor such as his facial expression and gaze. It also may include information about movement that is not available in the animations. We also note that participants in the video groups produced coarser units than those in the two animation groups. Given that coarse segmentation is associated with weaker correlations between segmentation and movement, it is possible that the video group's weaker correlations were caused by their coarser segmentation. In future work it will be important to explore this further, for example by constraining segmentation grain more tightly.

This experiment gave a surprising answer to our final question: Does one's prior conceptual representation of an activity affect the ongoing processing of movement information? If conceptual information such as that provided by an event schema allows viewers to establish a conceptual frame to integrate perceptual information during viewing, one would expect that providing a 40-s preview of an activity would be sufficient to establish such a frame in the animation-informed condition. If so, then the animation-informed groups should have differed from the animation-uninformed groups either in the strength with which movement features were related to segmentation, or in which particular features were correlated with segmentation. However, in neither experiment did we find evidence for such effects. Of course, this is a null result and as such it should be interpreted with caution. However, the differences between the animation groups and the video group indicate that this design had sufficient power to detect effects of the stimulus condition manipulation.

One account of this null result (suggested by an anonymous reviewer) is that for the impoverished animations, viewers might attend especially carefully to movement features, attempting to fit the movement information into whatever conceptual frame had been established. If so, then one would expect that activating a schema for an activity would preserve or even strengthen the degree to which movement features predict segmentation. However, on this interpretation one would expect that conceptual framing would affect *how* movement features relate to event segmentation, even if it does not affect *how strongly*. If this were the case one would expect that the preview manipulation would have affected the sign or magnitude of the correlations between individual movement features and segmentation; Fig. 6 shows this was clearly not the case.

Another possibility is that the 40-s previews provided in this experiment did not constrain viewers' conceptual representations enough to affect segmentation such that our measures could capture these effects. This could occur if viewers formed conceptual representations during the previews but then failed to retain them in memory throughout the animation. It also could occur if viewers' conceptual representations were not sufficient to tie the movement in the animations back to the objects and goals established in the preview. To investigate this, we fit linear models for the two animation groups using only the data from the first 40 s of each movie—the intervals that had previously been excluded from analysis because the animation-informed group previewed them prior to segmentation. If effects of conceptual framing on segmentation are present they should be especially strong for those first 40 s, because viewers will have just seen the previews and will have direct knowledge of how the movements in video relate to the objects in the scene. This analysis has the potential to overestimate effects of conceptual framing, because the animation-informed groups not only have been given a conceptual frame but also have seen the specific actions depicted as full-motion videos; this is why the first 40 s was excluded from the main analyses. Despite the potential to overestimate effects of conceptual framing, we observed no significant differences between the animation-informed and animation-uniformed conditions in the strength of the relationship between movement features and segmentation. There also were no significant differences in the correlations between individual movement features and fine segmentation. For coarse segmentation, three movement features showed small reductions in their correlations with segmentation that were significant at the .05 level but failed to survive corrections for multiple comparisons. In short, even in an analysis biased in favor of finding effects of conceptual framing on segmentation, there was little evidence for such effects.

4. General discussion

These experiments gave clear answers to three of the four questions we set out to answer. First, movement variables were significant predictors of event segmentation in naturalistic everyday activities. Second, when other sources of information for perceptual prediction were removed by reducing the live-action movies to simple animations, relations between movement and segmentation were strengthened. Third, relations between movement and segmentation were stronger for fine-grained segmentation than for coarse-grained segmentation.

These results are consistent with the account of event segmentation provided by EST (Zacks et al., 2007), which holds that event boundaries are detected in virtue of processing unpredicted feature changes. However, correlations between movement changes and event segmentation also may be predicted by other theories of event segmentation. For example, one account of event segmentation holds that the behavior stream consists of a spike structure in which brief bursts of change in features monitored by an observer form "points of definition" (Newtson, Hairfield, Bloomingdale, & Cutino, 1987). A change in one or more movement features could be defined to be such a burst. However, what determines which feature changes form spikes and what it means to be a point of definition are underspecified in this account. To discriminate between prediction error-based accounts and other ones it would be valuable to obtain more direct measures of prediction and prediction accuracy over time. This could be done using behavioral paradigms with an explicit predictive response, such as a visuomotor tracking task. Prediction errors also could be investigated without an explicit task by measuring physiological correlates such as electroencephalographic responses associated with errors (Holroyd & Coles, 2002).

The finding that movement features were more strongly correlated with fine segmentation than with coarse segmentation replicates previous results (Hard et al., 2006; Zacks, 2004). This pattern supports the proposal that fine segmentation depends more strongly than coarse segmentation on the processing of movement features, whereas coarse segmentation may depend more strongly on conceptual features (Zacks & Tversky, 2001). However, the fact that coarse segmentation of animations was not affected by the manipulation of prior information (Experiment 2b) offers a hint about the sort of information that may drive coarse segmentation: It suggests that the features driving coarse segmentation may not be particularly susceptible to top-down influences. In future research it would be valuable to test this hypothesis using other manipulations of top-down processing.

The finding that movement features predict when viewers will segment activity may have applications in the design of information systems. In domains such as medical data analysis (Christoffersen, Woods, & Blike, 2007), video surveillance (Chellappa, Cuntoor, Joo, Subrahmanian, & Turaga, 2008) and film production (Murch, 2001) it is important to segment a continuous data stream into psychologically meaningful events. Research in artificial intelligence and human-computer interaction has addressed this problem with some success (Chellappa et al., 2008; Davis, 1995; Rui & Anandan, 2000). Systems that use movement cues to define boundaries in data streams in the same way that people use those cues may produce more human-usable segmentations. Of course, in many domains it is impractical to invasively record actors' body movements as was done here. However, improvements in computer vision algorithms for biological motion recovery may render invasive motion tracking unnecessary (Sidenbladh, Black, & Fleet, 2000; Zhang & Troje, 2007). Automatic segmentation of activity into psychologically meaningful units would be helpful for producing visual summaries of data streams such as storyboards. Automatic segmentation also may be valuable as preprocessing for models designed to recognize actors, actions, or affect based on movement patterns (Pollick, Lestou, Ryu, & Cho, 2002; Troje, Westhoff, & Lavrov, 2005).

The final question addressed by these experiments was: Does one's prior conceptual representation of an ongoing activity affect the ongoing processing of movement information? We saw little evidence for effects of one's prior conceptual representation on movement processing. It is possible that we lacked detection power or that our analyses failed to quantify the movement variables affected by the experimental manipulation, or it could reflect a true null result. If so, it is at first blush a somewhat surprising result. In studies of the comprehension of texts (Bransford, Barclay, & Franks, 1972) and movies (Massad et al., 1979), providing a prior conceptual frame has consistently been found to have large effects on comprehension and later memory. The standard account of such phenomena is that providing a conceptual frame, say, by giving an informative title before reading a story, allows the reader to activate semantic knowledge structures that facilitate integration of incoming information. These knowledge structures, usually referred to as *event schemata*, capture information about actors, actions, objects, and their relations. One possibility is that typical event schemata represent information at a temporal grain coarser even than the coarse grain studied in Experiment 2b. The events in that study corresponded approximately to actions at the level of "spreading peanut butter on bread" or "folding a pile of shirts." It is possible that effects of conceptual framing on segmentation would be observed not at the level at which "spreading peanut butter" fits into "making a peanut butter sandwich" but at the level at which "making a peanut butter sandwich" fits into "packing lunch." However, this seems unlikely on its face. The temporal scale of the coarse grain measured in this study corresponds well to some of the components of event schemata measured in normative studies (Galambos, 1983; Rosen, Caplan, Sheesley, Rodriguez, & Grafman, 2003), and to units that have been identified with goal units and causal units in understanding everyday human behavior (Bower, 1982; Cooper & Shallice, 2006; Magliano, Taylor, & Kim, 2005; Trabasso & Stein, 1994; Zwaan & Radvansky, 1998). Another possibility is that conceptual framing affects how perceptual details are consolidated into long term memory or affects the reconstruction of events during retrieval without affecting the segmentation of activity into events on line. Effects of schemata on event memory sometimes reflect differences in reconstruction at retrieval time rather than effects on encoding processes (Anderson & Pichert, 1978; Thorndyke & Hayes-Roth, 1979). The present failure to find effects of prior conceptualization on event segmentation is consistent with such accounts. Whether one of these interpretations should be adopted awaits confirmation and extension of this intriguing null result.

Heraclitus wrote that "you can never step in the same river twice," in part because the river is no longer the same. This is an apt metaphor for the dynamic, fluid, and ever-changing stream of behavior. Given the complexity and variety of everyday activity and the fact that no previous event ever repeats perfectly, humans' ability to navigate our environment is really quite impressive. One way people cope with the complexity and dynamics of everyday activity is by segmenting it into meaningful chunks. The fact that some sequences of movement features are predictable appears to allow perceivers to pick out currents in the larger stream that form meaningful events.

Acknowledgements

This research was supported by the National Science Foundation (0236651). The authors would like to thank Becky Hedden and Jeff Stevenson for assistance carrying out the experiments.

References

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*(1), 1–12.

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: Lawrence Erlbaum Associates.

Barker, R. G., & Wright, H. F. (1954). Midwest and its children: The psychological ecology of an American town. Evanston, Illinois: Row, Peterson and Company.

Bower, G. (1982). Plans and goals in understanding episodes. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 2–15). Amsterdam: North-Holland Publishing Company.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193–209.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726.

Chellappa, R., Cuntoor, N. P., Joo, S.-W., Subrahmanian, V. S., & Turaga, P. (2008). Computational vision approaches for event modeling. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 279–341). Oxford: Oxford University Press.

Christoffersen, K., Woods, D. D., & Blike, G. T. (2007). Discovering the events expert practitioners extract from dynamic data streams: The mUMP technique. *Cognition, Technology, and Work, 9*, 81–98.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25*(2), 257–271.

Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review, 113*(4), 887–916 [discussion 917-31].

Davis, M. E. (1995). Media streams: Representing video for retrieval and repurposing.

Dickman, H. R. (1963). The perception of behavioral units. In R. G. Barker (Ed.), *The stream of behavior* (pp. 23–41). New York: Appleton-Century-Crofts.

Galambos, J. A. (1983). Normative studies of six characteristics of our knowledge of common activities. *Behavior Research Methods and Instrumentation, 15*(3), 327–340.

George, M. St., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain, 122*(Pt 7), 1317–1325.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale: L. Erlbaum.

Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General, 118*(2), 136–147.

Hanson, C., & Hirst, W. (1991). Recognizing differences in recognition tasks: A reply to Lassiter and Slaw. *Journal of Experimental Psychology: General, 120*(2), 211–212.

Hard, B. M., Recchia, G., & Tversky, B. (submitted for publication). The shape of action.

Hard, B. M., Tversky, B., & Lang, D. (2006). Making sense of abstract events: Building event schemas. *Memory and Cognition, 34*(6), 1221–1235.

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

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*, 201–211.

Lassiter, G. D. (1988). Behavior perception, affect, and memory. *Social Cognition, 6*(2), 150–176.

Lassiter, G. D., & Slaw, R. D. (1991). The unitization and memory of events. *Journal of Experimental Psychology: General, 120*(1), 80–82.

Lassiter, G. D., Stone, J. I., & Rogers, S. L. (1988). Memorial consequences of variation in behavior perception. *Journal of Experimental Social Psychology, 24*(3), 222–239.

Magliano, J. P., Taylor, H. A., & Kim, H. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory and Cognition, 33*(8), 1357–1367.

Maguire, E. A., Frith, C. D., & Morris, R. G. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain, 122*(Pt 10), 1839–1850.

Mann, R. & Jepson, A. D. (2002). Detection and classification of motion boundaries. In *Paper presented at the eighteenth national conference on artificial intelligence*. Edmonton, Canada.

Massad, C. M., Michael, H., & Newtson, D. (1979). Selective perception of events. *Journal of Experimental Social Psychology, 15*(6), 513–532.

Murch, W. (2001). *In the blink of an eye: A perspective on film editing* (2nd ed.). Los Angeles: Silman-James Press.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

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: Lawrence Erlbaum Associates.

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

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*(12), 847–862.

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. *Social Cognition, 5*(3), 191–237.

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research, 42*(20), 2345–2355.

Rosen, V. M., Caplan, L., Sheesley, L., Rodriguez, R., & Grafman, J. (2003). An examination of daily activities and their scripts across the adult lifespan. *Behavioral Research Methods and Computers, 35*(1), 32–48.

Rubin, J. M. & Richards, W. A. (1985). Boundaries of visual motion. AIM-835.

Rui, Y. & Anandan, P. (2000). Segmenting visual actions based on spatio-temporal motion patterns.In *Paper presented at the IEEE computer society conference on computer vision and pattern recognition*, Hilton Head, SC.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp. 33–58). Hillsdale, NJ: L. Erlbaum Associates.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23*, 473–500.
Sharp, R. M., Lee, J., & Donaldson, D. I. (2007). Electrophysiological correlates of event segmentation: How does the human mind process ongoing activity? *Paper presented at the Annual Meeting of the Cognitive Neuroscience Society*, New York.
Sidenbladh, H., Black, M. J., & Fleet, D. J. (2000). Stochastic tracking of 3D human figures using 2D image motion. In D. Vernon (Ed.), *Proceedings of the European conference on computer vision* (pp. 702–718). Berlin: Springer Verlag.
Speer, N. K., Reynolds, J. R., & Zacks, J. M. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*(5), 449–455.
Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53*, 125–140.
Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*, 236–257.
Thibadeau, R. (1986). Artificial perception of actions. *Cognitive Science, 10*(2), 117–149.
Thorndyke, P. W., & Hayes-Roth, B. (1979). The use of schemata in the acquisition and transfer of knowledge. *Cognitive Psychology, 11*(1), 82–106.
Trabasso, T., & Stein, N. L. (1994). Using goal-plan knowledge to merge the past with the present and the future in narrating events on line The John D. and Catherine T. MacArthur Foundation series on mental health and development. In M. Haith (Ed.), *The development of future-oriented processes* (pp. 323–349). Chicago: University of Chicago Press.
Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception and Psychophysics, 67*(4), 667–675.
Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory and Cognition, 28*(6), 1011–1021.
Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*(6), 979–1008.
Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001a). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651–655.
Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Situation changes predict the perception of event boundaries, reading time, and perceived predictability in narrative comprehension. *Journal of Experimental Psychology: General, 138*, 307–327.
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., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006a). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging, 21*(3), 466–482.
Zacks, J. M., Swallow, K. M., Speer, N. K., & Maley, C. J. (2006b). The human brain's response to change in cinema. *Abstracts of the Psychonomic Society, 11*, 9.
Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006c). Visual movement and the neural correlates of event perception. *Brain Research, 1076*(1), 150–162.
Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21.
Zacks, J. M., Tversky, B., & Iyer, G. (2001b). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29–58.
Zhang, Z., & Troje, N. F. (2007). 3D periodic human motion reconstruction from 2D motion sequences. *Neural Computation, 19*(5), 1400–1421.
Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185.