ELSEVIER

Cognitive Science 28 (2004) 979–1008

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Using movement and intentions to understand simple events

Jeffrey M. Zacks

Department of Psychology, Washington University, 1 Brookings Drive, Saint Louis, MO 63130, USA

Received 26 March 2004; received in revised form 3 June 2004; accepted 11 June 2004
Available online 25 September 2004

Abstract

In order to understand ongoing activity, observers segment it into meaningful temporal parts. Segmentation can be based on bottom-up processing of distinctive sensory characteristics, such as movement features. Segmentation may also be affected by top-down effects of knowledge structures, including information about actors' intentions. Three experiments investigated the role of movement features and intentions in perceptual event segmentation, using simple animations. In all conditions, movement features significantly predicted where participants segmented. This relationship was stronger when participants identified larger units than when they identified smaller units, and stronger when the animations were generated randomly than when they were generated by goal-directed human activity. This pattern suggests that bottom-up processing played an important role in segmentation of these stimuli, but that this was modulated by top-down influence of knowledge structures. To describe accurately how observers perceive ongoing activity, one must account for the effects of distinctive sensory characteristics, the effects of knowledge structures, and their interactions.

Keywords: Event perception; Movement; Knowledge structures

1. Introduction

One way people understand everyday activity is by segmenting it in time. A baseball game makes more sense if one understands that it consists of innings. A trip to the zoo might be remembered in terms of viewing different exhibits. People segment ongoing activity at salient boundaries, such as the ringing of a school bell, the arrival of a subway, the lights coming up in a theater, a swimmer reaching the shore. The result of this segmentation is that people perceive activity as consisting of *events*, each of which is a segment of time at a given location that is perceived to have a beginning and an end (Zacks & Tversky, 2001). The question addressed here is: On what basis do people identify the boundaries between meaningful event parts?

E-mail address: jzacks@artsci.wustl.edu.

0364-0213/$ – see front matter
doi:10.1016/j.cogsci.2004.06.003

There are two obvious candidates. First, a perceiver could break an activity into parts based on distinctive *sensory characteristics*. A school bell makes a recognizable sound, an arriving subway produces a blast of air, the lights in a theater dim before the movie. Dynamic *movement features* may be an especially powerful type of sensory characteristic. For example, when a swimmer in a triathlon reaches the shore and climbs onto dry land, the type and quantity of movement changes systematically. Any of these cues could directly cause one to perceive that a new event has begun, independent of any understanding of their significance for the larger activity. A dimming theater light might cause a person to perceive that an event has ended solely because it is a salient sensory change, even if that person had never been to the cinema and did not know that the film was about to begin. Seeing someone reach shore and climb from the water might be perceived as an event boundary by a person who did not know that this marked a new stage of the triathlon. This means of identifying event boundaries is *bottom-up*, because the sensory characteristics serve as direct cues to the structure of the event.

The second means by which an observer might identify event boundaries is by recognizing the activity in progress and parsing it in terms of one's knowledge of that activity. *Knowledge structures* are representations that capture recurring patterns of covariation. Knowledge structures for representing events have been described as schemas (e.g., Bartlett, 1932), scripts (Schank & Abelson, 1977), or situation models (e.g., Zwaan & Radvansky, 1998). For example, a rocket leaving the ground might be perceived as an event boundary because the viewer recognizes a rocket launch as beginning with a countdown, followed by take-off. One key feature of knowledge structures for events is that they represent actors' *intentions*. These may be of particular value in segmenting activity. For example, if a hiker came across a person chopping a tree, the hiker might recognize this as an instance of logging, and infer the goal of felling the tree. Therefore, the toppling of the tree might be perceived as an event boundary. Segmenting based on knowledge structures is *top-down*, because sensory characteristics are interpreted in terms of representations of the activity that: (a) are abstracted from the physical stimulus, and (b) depend on prior knowledge, i.e., knowledge that is present before the physical cue. This sort of top-down processing acts by modulating the bottom-up processing of sensory characteristics.

In short, event boundaries may be identified in a bottom-up fashion based on distinctive sensory characteristics, or in a top-down fashion based on knowledge structures. Movement features are an especially important sensory characteristic. Actors' intentions are a particularly important feature of knowledge structures for events.

1.1. Event segmentation correlates with both movement features and intentions

There is some evidence that both movement features and intentional structure correlate with observers' segmentation of events. This section briefly describes data suggesting that distinctive movement features correlate with perceptual event boundaries, and then describes data suggesting that actors' intentions are also correlated with event boundaries.

In one study (Newtson, Engquist, & Bois, 1977), observers watched short movies of single actors performing everyday activities, and marked boundaries between meaningful events by pressing a button whenever they believed a boundary occurred. Observers were either

given no special instructions, or were asked to segment the activity into either the smallest units they found meaningful (*fine* temporal grain) or the largest units they found meaningful (*coarse* temporal grain). The experimenters then coded the movies using choreographic notation, which provided a discrete coding of the position of the actor's body at one-second intervals. For each pair of successive intervals, they calculated the number of position features that changed. Transitions into and out of event boundaries had larger numbers of feature changes than periods that were not identified as event boundaries. This effect was strongest for the fine boundaries, and not statistically reliable for the coarse boundaries. This shows a correspondence between movement and the perception of event boundaries, particularly for fine-grained segmentation.

A similar correspondence between movement and event boundaries has been shown for longer events, though in this case the movements were much larger (Magliano, Miller, & Zwaan, 2001). In this study, participants watched a full-length feature movie and paused the movie whenever they felt the circumstances or situation in the film changed. (For example, in a James Bond film the setting might change from a space shuttle to an airplane.) The researchers coded each transition between successive shots in the movie to identify those points at which the spatial location changed. (A shot is a continuous sequence as observed from a single camera. Mean shot durations ranged from 3.95 s to 6.18 s across movies.) Boundaries tended to be near transitions where the spatial location changed, suggesting that movement of the action from one location to another was associated with the perception of an event boundary.[1]

Neurophysiological studies converge with these data in supporting a role for movement features in event perception. Neuroimaging evidence indicates that, when observers watch movies of everyday activity, brain regions involved in motion processing transiently increase in activity at those moments in time that the observers later indicate were perceived as event boundaries (Zacks, Braver et al., 2001). In particular, the MT complex, an extrastriate visual area that contains cells selective for direction and speed of motion, was strongly activated (Speer, Swallow, & Zacks, 2003). At first blush this might appear to militate for the view that bottom-up processing of motion information drives event perception. However, this conclusion is not warranted because the MT complex is known to be modulated by recurrent projections from parietal and frontal cortex in response to changes in attention and task demands (e.g., Friston & Buchel, 2000).

There is also evidence that actors' intentions correlate with how events are perceived. Wilder (1978a, 1978b) showed that viewers segment activity into smaller units when they cannot predict the actors' goals. Participants viewed movies of an actor performing a goal-directed activity. When the activity was predictable, they segmented it into relatively coarse-grained units. Segmentation was reliably finer in grain when the activity either became unpredictable after being initially predictable (Wilder, 1978a), or was unpredictable in terms of both moment-to-moment relations and larger structure (Wilder, 1978b). One possibility is that observers segment activity in terms of the largest goals they can identify, and when things become unpredictable those goals are more fine-grained.

A study of infant perception examined the moment-by-moment effect of goals on perception. Infants 10–11 months old watched short sequences taken from a movie depicting a woman cleaning her kitchen (Baldwin, Baird, Saylor, & Clark, 2001). After familiarization with a sequence, the infants were shown the same sequence with a pause inserted either at the moment

the actor achieved a goal, or just before or after. The infants looked longer at the sequences with pauses placed away from the goal achievements, suggesting that their natural parsing of the activity placed boundaries at those points at which a goal was achieved.

1.2. A simple model

The available data indicate that both movement features and intentions are correlated with event segmentation. Further, they suggest that the fine-grained segmentation and coarse-grained segmentation differ, with fine-grained segmentation being more directly dependent on movement features. One possibility, suggested in particular by Wilder's (1978a, 1978b) results, is that finer-grained boundaries are determined by bottom-up processing of sensory characteristics, but how these fine-grained segments group into coarse-grained segments is modulated by top-down processing based on prior knowledge, including intentions (Zacks & Tversky, 2001). This implies that the relationship between movement features and the perception of event boundaries will be strongest for fine-grained event parts, whereas the effect of inferences about intentions will be stronger at coarser grains.

These considerations led to the development of a model governing the role of movement features in event segmentation. A graphical representation of the model is given in Fig. 1. The model proposes that sensory characteristics are processed in a bottom-up fashion on an ongoing basis, providing input to feature detectors, in turn leading to the perception of event segments (among other consequences). This processing stream is modulated by the influence of knowledge structures, providing top-down processing such that identical sensory cues can lead to different patterns of segmentation depending on the current state of knowledge. (Knowledge structures may become activated due to configurations of sensory characteristics. They may also

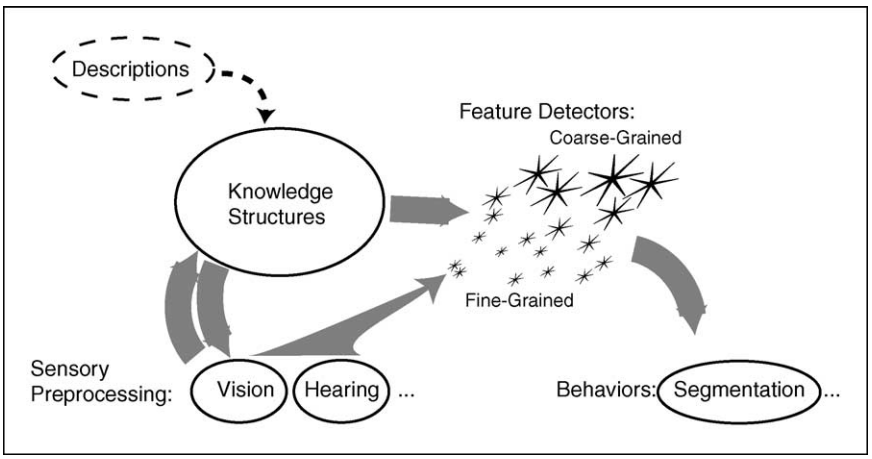

Fig. 1. A model of the role of movement features in event segmentation. A bottom-up processing stream computes event segments from sensory characteristics, using feature detectors that vary in their temporal grain. This processing stream interacts bidirectionally with knowledge structures, which allow attributes such as actors' intentions to influence processing. Knowledge structures have their strongest effects on coarse-grained feature detectors. Dashed lines indicate that the mechanism by which explicit descriptions of the activities (i.e., the interpretation manipulations) are interpreted and activate knowledge structures is outside the scope of the model.

be activated in other ways, for example, by verbal descriptions. However, these mechanisms are outside the scope of the model.) Feature detectors are sensitive to a range of temporal grains, and knowledge structures have most effect on those feature detectors with coarse temporal tuning.

The implications of the model can be expressed as four postulates:

1) Movement features contribute to the identification of fine event segments.
2) Grouping these fine segments into larger units can be based on aspects of the activity other than movement features. Observers rely less on movement features as the grain of encoding becomes larger.
3) Inferences about actors' intentions can affect how and to what extent movement features drive the identification of event segments.
4) Inferences about actors' intentions can be influenced by both intrinsic features of the stimulus and by top-down information.

Regarding the third postulate, note that the model predicts that top-down information can have two different effects on processing. First, it can affect *to what extent* movement features drive event segmentation, because the output of knowledge structures is combined with the output of sensory processing in the input to the feature detectors. Second, it can affect *how* movement features drive segmentation, by the recurrent connections between knowledge structures and sensory processing.

1.3. How do movement and intentions causally drive event perception?

The data described above are consistent with the possibility that both distinctive sensory characteristics (in particular, movement features) and knowledge of an activity (in particular, actors' intentions) cause viewers to perceive event boundaries. However, there is a problem with this conclusion: In naturally occurring events, sensory characteristics and knowledge structures are confounded (Zacks & Tversky, 2001). Consider some of the previous examples. Distinctive sensory characteristics correspond with knowledge structures: A dimming theater light is a distinctive sensory characteristic, but also is a reliable indicator that the projectionist has achieved the goal of waiting for the patrons to arrive and has adopted the goal of starting the movie. A triathlete reaching shore is physically distinctive, but also a reliable indicator (at least to other triathletes) that the athlete has finished the first of the three legs of the race, and is about to begin the bicycling leg. Conversely, knowledge structures correspond with distinctive sensory characteristics: The toppling of a tree or the launching of a rocket not only correspond with information in knowledge structures, but also produce distinctive configurations of sensory characteristics (movement and sound). Previous studies have measured either movement features or actors' intentions, but none have manipulated the two independently. Thus, it is possible that apparent effects of movement features on perceived event structure are actually due to covarying changes in actors' intentions, or that the previously reported tendency to segment activity at goals is actually due to covarying movement features. Thus, two important questions remain. First, to what degree do movement features determine how viewers segment events (postulates 1 and 2 in the model)? Second, does knowledge of actors' intentions modulate the role movement features play in segmenting events (postulates 3 and 4)?

To better characterize the effect of movement features and intentions on viewers' perception of event boundaries, the experiments reported here manipulated the presence and timing of distinctive movement features, and also manipulated viewers' inferences about the goals present in the activity. Participants viewed simple animations that could be interpreted either as intentional action or as random motion, and segmented these animations into fine or coarse segments. Because the animations were simple and rendered by a computer, it was possible to precisely characterize the movement information in each movie. This sort of stimulus has been used productively to study the perception of causality and intention (e.g., Bassili, 1976; Heider & Simmel, 1944). To summarize the outcome: The results of three experiments indicated that movement features do help determine where viewers perceive event boundaries (postulate 1), more for fine-grained than coarse-grained events (postulate 2). The data also suggested a role for intentions in modulating the processing of movement features (postulate 3), even in these simple stimuli. Finally, viewers' attributions of intentions to the stimuli depended both on intrinsic features of the stimuli and on top-down information provided by experimental instructions (postulate 4).

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty Washington University students (8 male, ages 18–22) participated in partial fulfillment of a course requirement. None of the participants had participated in a previous study of event segmentation in our laboratory. An additional five were replaced due to equipment failure, and one was replaced because she identified nearly as many coarse boundaries (126) as fine boundaries (131).

2.1.2. Materials

A 600 s animation depicted the movements of an orange circle and a green square within a square region with a white background. The objects moved randomly with momentum approximating random forces acting in the presence of friction. This configuration was chosen for two reasons. First, the presence of two objects provided information about the relative movement of the objects, which may be particularly important for event perception. Second, previous research had demonstrated that people can accurately interpret the movements in such stimuli when they are generated by human actors (Blythe, Todd, & Miller, 1999).

The movie was rendered on a 480 pixel square canvas, and the two objects were each 20 pixels in diameter. Coordinates were calculated based on a unit square and then multiplied by 480 for rendering. The movie canvas was depicted in the center of a 832 $\times$ 624 pixel monitor with a white background, such that there was no visual marking of the boundary of the canvas.

The movie began with the objects placed at the horizontal center of the canvas, with the green square .1 units (48 pixels) above the vertical center and the orange circle .1 units (48 pixels)

below. On each frame, the positions of the objects were updated according to the following equations:

$$x_i = x_{i-1} + D\partial_{x,I-1} + Ar_{x,I}(1 - D) \quad (1a)$$

$$y_i = y_{i-1} + D\partial_{y,I-1} + Ar_{y,I}(1 - D) \quad (1b)$$

The variables x and y denote the x and y position of the object, and the subscript i denotes the current time step. The value $\partial_{x,I-1}$ is the amount the object moved in the x direction on the previous time step, i.e., $\partial_{x,I-1} = x_{i-1} - x_{i-2}$. Similarly, $\partial_{y,I-1} = y_{i-1} - y_{i-2}$. The variables $r_{x,I}$ and $r_{y,I}$ are random numbers drawn from the uniform distribution $[-.5, .5]$. The parameters D and A control the relative contributions of previous velocity and noise to the object's movement.

Two constraints were placed on the objects' movement. In order to assure that the objects remained on the canvas, any updates that placed an object within .05 units of the edge of the screen were rejected and recalculated. (However, the ∂ values were retained.) To avoid the objects passing over each other, movements that placed the objects within .1 of each other were also rejected and recalculated.

For this experiment, the animation parameters were set to $A = .2$ and $D = .975$. Movies were rendered at 10 frames/s, which was adequate to provide a percept of smooth motion with these simple objects. This resulted in an animation that could be described as relatively leisurely in pace and smooth in movement trajectory. An illustration of the movement trajectories is given in Fig. 2; the full movie is available in the Cognitive Science on-line annex (see Movie 1 in the Cognitive Science on-line annex at http://www.cognitivesciencesociety.org/supplements/).

The movies were presented and responses were collected using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Power Macintosh computer (Apple, Cupertino CA). The stimulus on screen measured 20 cm and was viewed from approximately 45 cm, for a viewing angle of approximately 25°.

2.1.3. Procedure

After providing informed consent, participants were told that they would be viewing a short animation. Each participant was randomly assigned to one of two interpretation conditions. Those in the intentional group were told that the movie depicted the recorded movements of two people in a room completing a goal-directed activity. Those in the random group were told that the movie was randomly generated. Both groups were asked to divide the movie into natural and meaningful units while watching it, by pressing the button each time one meaningful unit ended and another began. Equal numbers of participants were tested in each group.

The *grain* of segmentation was manipulated within participants. Within each of the two interpretation groups, half of the participants were asked to segment the activity during the first viewing into the smallest units that were natural and meaningful to them (*fine* units), and half were asked to produce the largest units that were natural and meaningful (*coarse* units). It was emphasized that there was no right or wrong answer, and that they should report where they perceived the natural segment boundaries to lie. During the second viewing, each participant was instructed to segment the movie at the grain not tested on the first viewing.

After participants read these instructions and the experimenter answered any questions, participants viewed a 120 s practice stimulus, generated using the same parameters as the

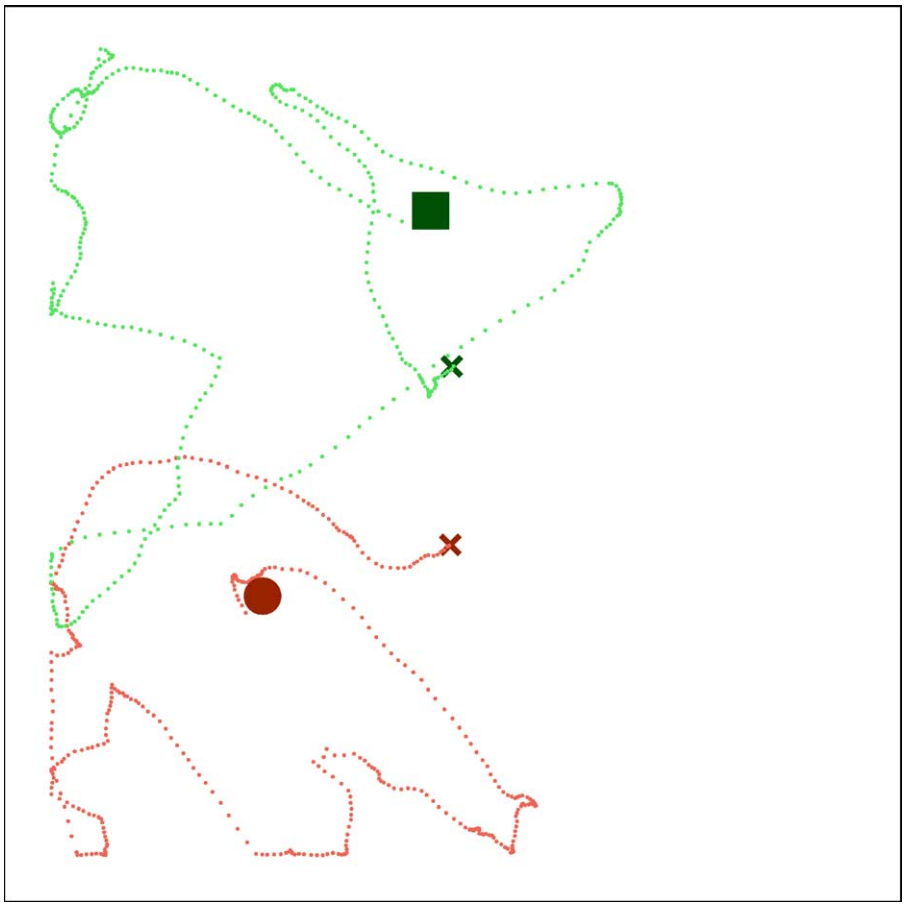

Fig. 2. Illustration of the animations. The image shows the two objects depicted in the animation, a green square and an orange circle. The small dots depict the path of each of the objects during the first 40 s of the animation used in Experiment 1, sampled at 10 frames/s, and *x*'s mark the objects' initial positions. (The black frame bordering the canvas was not visible to the participants.)

test stimulus. They were then asked if they understood the task and given the opportunity to ask more questions. Then, each participant segmented the test movie twice, once at each segmentation grain.

After performing the segmentation task, participants were asked to complete a questionnaire asking them about their performance. The first question asked for a description in their own words of what had happened in the movie. The second and third questions asked about each of the individual objects. The fourth and fifth questions asked the participants to rate how *goal-directed* and how *random* the movements of the objects appeared to be on five point Likert scales. The sixth and seventh questions asked the participants to describe how they decided where the small and large unit boundaries were. Finally, participants were asked if there was anything else that might be of interest.

2.2. Results

All analyses reported here were conducted by discretizing time in the movie into one-second bins (Zacks, Tversky, & Iyer, 2001). For each viewer, a bin was considered a *breakpoint* during coarse or fine segmentation if the viewer pressed the key during that one-second interval. For analysis at the group level, one can calculate a *breakpoint histogram*, i.e., a plot of the number of viewers who identified a given time bin as a breakpoint.

2.2.1. Unit size

A logical first question about participants' event segmentation is: How many units did they identify? (This can also be phrased as, How big were the units?) Counts of the number of units identified on each viewing and the mean length of units were calculated for both groups, for fine-grained and coarse-grained segmentation, and are summarized in Table 1. An analysis of variance (ANOVA) was conducted on the number of breakpoints during each viewing, with interpretation condition as a between-participants variable and grain as a repeated measure. (Breakpoint counts were used because the distributions of lengths were positively skewed.) As expected, when participants were asked to segment into fine units they produced more breakpoints (smaller units), $F(1, 18) = 58.2$, $p < .001$. Table 1 also indicates that participants given the random motion interpretation produced substantially more breakpoints (smaller units) than those given the intentional motion interpretation, $F(1, 18) = 5.9$, $p = .03$. The interaction between group and segmentation grain was not reliable, $F(1, 18) = 2.5$, $p = .13$.

2.2.2. Role of movement features

Because the objects in the animations were effectively points moving in the plane, movement features could easily be characterized in terms of their position and aspects of the higher moments of position. Movement features were calculated for each frame of the animation and averaged over one-second intervals for comparison with the behavioral segmentation data. The general approach was to consider a large number of features, and use stepwise multiple regression to identify those features that best predicted where viewers would segment the movie into coarse and fine units, for each of the two interpretation conditions. The features selected capture the absolute and relative positions of the objects and their first two higher moments (velocity, acceleration). In addition, maxima and minima of the higher moments

Table 1
Mean breakpoint counts and unit sizes as a function of grain of segmentation and interpretation condition in Experiment 1

Grain	Interpretation	Number of breakpoints	Unit length
Fine	Random	110.3 (47.9)	6.29 s (2.52 s)
	Intentional	69.6 (22.8)	9.95 s (5.57 s)
Coarse	Random	56.3 (33.9)	18.2 s (20.7 s)
	Intentional	34.2 (8.32)	17.9 s (3.97 s)

Standard deviations in parentheses.

Table 2
Movement features analyzed in Experiments 1–3

Feature	Description
Position (4)	Planar *x* and *y* location, for each of the two objects
Velocity (4)	Velocity in *x* and *y* for each of the two objects (calculated by numerical differentiation of position)
Acceleration (4)	Acceleration in *x* and *y* for each of the two objects (calculated by numerical differentiation of velocity)
Speed (2)	The speed of each object (i.e., the magnitude of the object's instantaneous velocity)
Acceleration magnitude (2)	The magnitude of each object's instantaneous acceleration
Relative Position (2)	Relative *x* and *y* location of the two objects
Distance (1)	The distance between the two objects
Relative Speed (1)	How fast the objects were moving toward or away from each other (calculated by numerical differentiation of Distance)
Relative Acceleration (1)	How fast the objects were accelerating toward or away from each other (calculated by numerical differentiation of Relative Speed)
Speed Maxima (2)	Binary features indicating whether each of the objects was at a local maximum in speed
Speed Minima (2)	Binary features indicating whether each of the objects was at a local minimum in speed
Acceleration Maxima (2)	Binary features indicating whether each of the objects was at a local maximum in acceleration
Acceleration Minima (2)	Binary features indicating whether each of the objects was at a local minimum in acceleration
Distance Maxima (1)	A binary feature indicating whether the distance between the objects was at a local maximum
Distance Minima (1)	A binary feature indicating whether the distance between the objects was at a local minimum
Relative Acceleration Maxima (1)	A binary feature indicating whether the relative acceleration of the objects was at a local maximum
Relative Acceleration Minima (1)	A binary feature indicating whether the relative acceleration of the objects was at a local minimum

Numbers in parentheses describe how many scalar values were required to code each feature (e.g., position requires four scalars to encode the *x* and *y* location of both objects).

were calculated because extrema are important to perception in a range of situations (e.g., Hoffman & Richards, 1984). The features used in all experiments are listed in Table 2.

What if the effect of a movement feature on segmentation is not instantaneous, but instead includes some anticipation or lag? For example, suppose that viewers segmented the activity by looking for local minima in position (i.e., moments at which the two objects came closest to touching). An observer cannot be sure that a local minimum has been reached until after it is over. Furthermore, there may be some delay associated with deciding that a given moment is indeed a local minimum. If this were the case, the correlation between this movement feature and event segmentation would be maximal only if two time series were shifted slightly relative to each other. However, it cannot be said a priori what shift is optimal, or that different movement features necessarily have the same optimal shift. It cannot even be pre-determined

whether to shift the segmentation data forward or backward relative to a given movement feature. This may seem counterintuitive, but anticipatory couplings between segmentation and movement features cannot be ruled out, because autocorrelation in the movement data makes it possible that participants may react to a feature in the movement data, which is then followed consistently by another pattern. Therefore, optimal shifts were estimated separately for each movement feature prior to the primary analysis. For each movement feature, the cross-correlation between that feature and the overall likelihood of segmentation was calculated for lags between −5 and 5 s. (The overall segmentation likelihood was calculated by summing the breakpoint histograms for each combination of grain and interpretation condition, after dividing these by their standard deviations to control for the fact that each histogram was based on a different number of breakpoints.)

Linear regression was used to calculate the proportion of variance in event segmentation accounted for by movement features. Forward stepwise regression was used to estimate which of the shifted movement features predicted where viewers segmented. Separate analyses were computed for each combination of grain and interpretation condition. For each analysis, the total amount of variance accounted for by movement features, and the first variables to enter the regression, are reported in Table 3. As can be seen in the table, movement features accounted for significant variance in all conditions. Movement features predicted the least variance when participants were segmenting at a coarse grain and had been told that the movie depicted intentional activity; however, none of the four conditions differed significantly from each other (as tested using Fisher's *z* transformation).

To the extent that movement features were predictive of segmentation, the features that accounted for the most variance were related to the distance between the objects and to the objects' acceleration. The acceleration of the two objects was correlated ($r = .63$);[2] thus, once one of the two objects' acceleration magnitude was entered into the regression equation, the other was superfluous. Therefore, the fact that the orange circle's magnitude of acceleration

Table 3
Relationship between movement features and segmentation in Experiment 1

Grain	Interpretation	R^2 (95% confidence intervals)	Most predictive features[a]
Fine	Random	.14 (.09–.20)	−Distance, +Distance Minima, +Acceleration$_S$, −Speed Minima$_S$
	Intentional	.16 (.11–.22)	+Acceleration Magnitude$_C$, +Distance Minima, −Relative Speed, +Speed$_C$
Coarse	Random	.16 (.11–.22)	+Distance Maxima, +Speed$_C$, +Acceleration$_S$, +Relative Acceleration, +Speed$_S$, −Acceleration$_{CY}$
	Intentional	.11 (.06–.16)	+Speed$_C$, −Relative Speed

For each grain of segmentation and interpretation condition, the table gives the proportion of variance accounted for by movement features (R^2). The rightmost column lists the movement features that accounted for most of the variance, with a sign (+ or −) to indicate whether the correlation between that variable and segmentation was positive or negative.

[a] Features are listed in the order they entered the stepwise regression, up to the last feature that accounted for at least 1% of incremental variance. Abbreviations in subscripts refer to one of the two objects ("C" for circle, or "S" for square) or to one of the two dimensions of movement ("*X*" or "*Y*").

entered the regression for fine segmentation under intentional instructions, whereas the green square's acceleration magnitude appears to be important for the other conditions, should not be taken as significant (see Table 3).

2.2.3. Agreement across conditions

Agreement between the two groups was evaluated by calculating breakpoint histograms for each group for fine-grained and coarse-grained segmentation, and comparing those across the groups. The two groups showed agreement about the locations of fine-grained boundaries, $r = .32$, $df = 598$, $p < .001$. At a coarse temporal grain the agreement was reduced but still statistically significant, $r = .19$, $df = 598$, $p < .001$. The reduction in agreement was statistically reliable as tested by Fisher's z transformation, $z = 2.32$, $p = .01$. In short, the two groups agreed better about the location of fine-grained boundaries than about coarse-grained boundaries.

2.2.4. Questionnaire responses

Participants rated the level of goal-directedness and randomness of the stimulus using a seven point Likert scale. The groups were compared for these ratings with t-tests. As expected, the intentional interpretation group rated the movie as more goal-directed ($M = 4.20$, $SD = 1.23$) than the random interpretation group ($M = 3.30$, $SD = 1.57$). However, this difference was not statistically reliable, $t(18) = 1.43$, $p = .17$. Also as expected, the intentional interpretation group rated the movie as less random ($M = 4.10$, $SD = 1.60$) than the random interpretation group ($M = 4.50$, $SD = 1.72$). This difference also was not reliable, $t(18) = .54$, $p = .60$. Although the small sample sizes placed limits on the power to detect differences between the groups, the failure of the manipulation to produce reliable differences on the Likert scale measures suggests that the manipulation was weak. In this experiment the number of participants was too small to permit a formal analysis of the movie descriptions; these will be reported for Experiments 2 and 3.

In this experiment and the two that followed, the strategies for segmentation were coded by two independent raters. This coding suggested some relationships between the experimental manipulations and the strategies reported, but the only statistically reliable effect was that participants given intentional interpretations were more likely to report looking for intentional features when segmenting. This is unsurprising, and could reflect in part demand characteristics. Therefore, the strategy reports will not be described in detail.

2.3. Discussion

When participants segmented this simple animation into fine-grained units, 14–16% of their unit boundary placement was accounted for by simple features of object motion. This was true independent of whether they were told the movie depicted intentional activity or random motion. While not overwhelming, this relationship indicates a clear role for movement features in the segmentation of activity into fine-grained units.

When participants segmented the movie into coarse-grained units and were told the movie depicted the intentional activity of two people, the relationship between movement features and event segmentation was (nonsignificantly) weaker. This suggests that people rely more on abstract emergent features to group fine-grained units into larger structures, at least if given grounds for doing so.

Agreement between participants given random and intentional cover stories also differed across temporal grain. The two groups showed modest agreement regarding the location of fine-grained boundaries, and weaker agreement at a coarse grain. This was true despite the fact that the random interpretation group's fine units were considerably "finer" than those of the intentional interpretation group. Such a result is consistent with the view that both groups segmented based on movement features at a fine grain, but at a coarse grain the intentional interpretation group tended to rely on other aspects of the stimulus.

The data from Experiment 1 indicated some undesirable psychometric properties. First, the two groups differed in the size of their coarse and fine units. While this result itself indicates the two groups segmented the movie differently, it means that the group comparisons of movement features and level of agreement were not comparing equivalent measures as tightly as would be desired. Second, the manipulation checks provided only weak evidence that the cover stories actually influenced how the participants interpreted the movie. Finally, the small sample size limited the power of the movement feature and group difference analyses.

Experiment 2 was conducted to extend these results while: (a) better controlling the size of participants' coarse and fine units, (b) strengthening the interpretation manipulation, and (c) testing a larger sample.

3. Experiment 2

3.1. Method

3.1.1. Participants

Forty-eight Washington University students (34 female, ages 17–22) participated in partial fulfillment of a course requirement. None of the participants had participated in a previous study of event segmentation in our laboratory.

3.1.2. Materials

The major modification in Experiment 2 was designed to strengthen the interpretation manipulation. As before, each participant was assigned to either a random interpretation or an intentional interpretation condition. Equal numbers were tested in each group. The instructions for the random group were as in Experiment 1: They were told that the animation was randomly generated. To strengthen the effectiveness of this instruction, they were shown the equations describing the objects' motion, and it was explained how the computer could use them to generate movement trajectories. The intentional group was told that the animation was generated by two people playing a video game, in which each person controlled one of the objects using the computer keyboard. To strengthen this interpretation, participants played a video game that corresponded to this description, similar to the one described by Blythe et al. (1999). Participants viewed a green square and an orange circle, exactly as for the target animation. Each participant used the arrow keys to control the square, while the experimenter controlled the circle. Participants were instructed that they should chase the experimenter's circle with their square. Each participant played the game for a few minutes.

A new animation was constructed using the same software and algorithms as for Experiment 1 (see Movie 2 in the Cognitive Science on-line annex at http://www.cognitivesciencesociety.org/supplements/). Two significant changes were made to the parameters in order to generate an animation that would resemble the movements of two objects controlled by people playing a video game. First, the constraint that prohibited the objects from colliding was removed. Second, a new set of parameters for the movement update equations was selected to produce faster, jerkier trajectories. The parameters chosen were $A = 1.0$ and $D = .9425$. These were derived from data collected by Blythe et al. (1999), whose participants played a video game similar to the one used in the present study. The selected values produced trajectories with mean velocity and mean acceleration that were in the middle of the distribution of velocity and acceleration for the Blythe et al. stimuli. A 600 s target movie was generated, as well as a 120 s practice movie, just as in Experiment 1.

The animation was presented using the same hardware and software setup as in Experiment 1.

3.1.3. Procedure

The procedure was similar to that of Experiment 1. After providing informed consent participants were briefed regarding the task and assigned to either the random or intentional interpretation conditions. They either were shown the object movement equations or played the video game, as described above. Each participant segmented the target animation at both a fine and coarse temporal grain, with order counterbalanced across participants. To reduce individual differences in the temporal grain of coarse and fine segmentation, a more elaborate practice procedure was used. To ensure that the segmentation grain for coarse and fine unit lengths were different from each other, and consistent with viewers' natural grain of segmentation of everyday activity, target ranges were selected based on previous data for segmentation of naturalistic live action movies (Zacks, Tversky, & Iyer, 2001). Intervals were chosen surrounding the median unit lengths in those data, and were: 4–5.5 fine-unit breakpoints per minute, and 1–2.5 coarse-unit breakpoints per minute. Each participant was given the opportunity to segment the 120 s practice movie and their breakpoint count was compared to these ranges. If the number of breakpoints produced was outside the appropriate range, the participant was given feedback and asked to segment the practice movie again. If the participant's segmentation was still not in the target range after the second viewing of the practice movie, feedback was given again and a third viewing was presented. If the participant did not reach target segmentation by this point, they were excused and replaced.

This procedure has the advantage of more precisely characterizing segmentation at fine and coarse grains, reducing extraneous variability and potential group differences in segmentation grain, without biasing where participants segment. Because the iterative feedback procedure is necessarily vague, some participants failed to reach the criterion before the cutoff. In Experiment 2, four participants were excluded because they failed to reach the unit size criterion. Two additional participants successfully completed the practice procedure, but then produced unusually few (3) or many (124) breakpoints during one viewing. Both were well outside the distributions for their experimental conditions, so their data were replaced. One participant failed to comply with experimental instructions, and two participants experienced equipment failure; their data were also replaced.

Table 4
Mean breakpoint counts and unit sizes as a function of grain of segmentation and interpretation condition in Experiment 2

Grain	Interpretation	Number of breakpoints	Unit length
Fine	Random	42.0 (11.9)	15.2 s (5.04 s)
	Intentional	46.8 (14.8)	14.7 s (8.55 s)
Coarse	Random	19.3 (6.4)	34.5 s (18.1 s)
	Intentional	20.5 (8.8)	34.8 s (19.1 s)

Standard deviations in parentheses.

Each participant completed the training procedure and then segmented the target movie for both coarse-grained and fine-grained segmentation, with order counterbalanced. After completing the second segmentation, each participant completed a questionnaire identical to that used in Experiment 1, and was then debriefed and excused.

3.2. Results

3.2.1. Unit size

The sizes of coarse and fine units for both groups are plotted in Table 4. As can be seen in the table, the training procedure succeeded in inducing participants to segment at a grain comparable to the target ranges previously observed in the segmentation of everyday activities (Zacks, Tversky, et al., 2001), and succeeded in controlling grain of segmentation across the groups. An ANOVA on the breakpoint counts with interpretation as a between-participants variable and segmentation grain as a repeated measure revealed no main effect of group, $F(1, 46) = 1.29$, $p = .26$, and no interaction between group and segmentation grain, $F(1, 46) = 1.06$, $p = .31$. The expected main effect of segmentation grain was highly reliable, $F(1, 46) = 197.1$, $p < .001$.

3.2.2. Role of movement features

Movement features were calculated and analyzed as for Experiment 1. Forward stepwise regressions were used to estimate which movement features predicted where viewers segmented the movie for each combination of interpretation condition and grain. For each regression, the first variables to enter the regression and the total amount of variance accounted for by movement features are reported in Table 5. Differences in the degree to which movement features accounted for variance in segmentation were assessed with Fisher's z transformation. Movement features accounted for significantly more variance in the fine-grained segmentation conditions than in coarse-grained segmentation conditions (all p's $< .001$). Movement features did not account for significantly different amounts of segmentation in the two interpretation groups at a fine grain ($p = .41$) or a coarse grain ($p = .40$).

For all experimental conditions, the distance between the two objects was the best predictor of perceptual segmentation. Acceleration of the square and relative acceleration of the two objects were also consistent predictors of segmentation. (As in Experiment 1, the two objects' acceleration magnitudes were correlated, $r = .47$, which explains why only the square's accel-

Table 5
Relationship between movement features and segmentation in Experiment 2

Grain	Interpretation	R^2 (95% confidence intervals)	Most predictive features[a]
Fine	Random	.32 (.26–.39)	−Distance, +Acceleration Magnitude$_S$, +Distance Maxima, +Relative Acceleration Maxima
	Intentional	.33 (.27–.40)	−Distance, +Acceleration Magnitude$_S$, −Relative Position$_X$, −Velocity$_{S,Y}$, −Distance Maxima, −Relative Acceleration Maxima
Coarse	Random	.17 (.12–.23)	−Distance, −Speed$_S$, −Distance Maxima, +Acceleration Maxima$_C$, +Acceleration Magnitude$_S$
	Intentional	.18 (.13–.24)	−Distance, −Distance Maxima, +Acceleration Magnitude$_S$, −Velocity$_{C,X}$, −Acceleration Minima$_C$

For each grain of segmentation and interpretation condition, the table gives the proportion of variance accounted for by movement features (R^2). The rightmost column lists the movement features that accounted for most of the variance, with a sign (+ or −) to indicate whether the correlation between that variable and segmentation was positive or negative.

[a] Features are listed in the order they entered the stepwise regression, up to the last feature that accounted for at least 1% of incremental variance. Abbreviations in subscripts refer to one of the two objects ("C" for circle, or "S" for square) or to one of the two dimensions of movement ("*X*" or "*Y*").

eration magnitude entered the regression equations.) For fine-grained segmentation, the two groups showed nearly identical relationships between movement features and segmentation. For coarse-grained segmentation there was more divergence in which features predicted where the two groups segmented.

3.2.3. Agreement across conditions

As for the previous experiment, agreement between the two groups for both fine-grained and coarse-grained segmentation was compared by correlating the two groups' breakpoint histograms. It should be expected that these correlations would be higher overall, because: (a) the samples were larger, and (b) the practice procedure was designed to reduce within-group variability and thus increase reliability. This was indeed the case: The two groups' breakpoint histograms correlated highly for both fine segmentation ($r = .86$, $p < .001$) and coarse segmentation ($r = .87$, $p < .001$). The two correlations did not differ reliably ($z = .49$, $p = .31$). In short, agreement across groups was higher in this experiment and, unlike Experiment 1, did not vary as a function of grain of segmentation.

3.2.4. Questionnaire responses

Results of the questionnaire Likert scales were analyzed as in Experiment 1. As expected, participants in the intentional interpretation group rated the movie as more goal-directed ($M = 3.96$, $SD = 1.33$) than those in the random interpretation group ($M = 2.67$, $SD = 1.24$). This difference was larger than in Experiment 1. This, plus the larger sample size, led to a reliable difference between the two groups, $t(46) = 3.47$, $p < .001$. Also as expected, participants in the random attribution group rated the movie as more random ($M = 5.17$, $SD = 1.31$) than those

in the intentional attribution group ($M = 4.46$, $SD = 1.50$). This difference was also larger than that in Experiment 1; however, it did not reach statistical significance, $t(46) = 1.74$, $p = .09$. Note that participants in the current experiment rated the movie as less goal-directed and more random overall, compared to Experiment 1.

The descriptions of the movies were coded by the author and a trained rater. Each rater first recorded whether each participant's description of the movie included words that clearly indicated an intention. Examples include "chase," "lead," "find," "try," "prevent," and "goal." The two raters agreed for 41 of 48 descriptions (85%); the rest were excluded from analysis. For those descriptions on which both coders agreed, 100% of the 22 participants in the intentional interpretation condition used intentional terms to describe the activity. However, only 37% (7 of 19) of the participants in the random interpretation condition used intentional terms. This group difference was statistically reliable by Fisher's exact test of goodness of fit, $p < .001$.

3.3. Discussion

The primary methodological goals of Experiment 2 were to better control participants' grain of segmentation across groups and to increase the degree to which participants interpreted the movie as intentional activity or random motion. The modifications made to the procedure were effective in this regard: The two groups did not differ in their grain of segmentation, and the manipulation checks showed that the attribution manipulation was more effective than that of Experiment 1.

In this modified design, movement features accounted for a substantial amount of the variance in both groups' segmentation of the movie—overall, more than in Experiment 1. As in Experiment 1, movement features predicted fine-grained segmentation better than coarse-grained segmentation. Unlike Experiment 1, movement features accounted for significant variance in coarse-grained segmentation for the intentional group (20%)—in fact, slightly more than for the random group (18%). Also unlike Experiment 1, the two groups showed strong agreement regarding the location of coarse event boundaries as well as fine boundaries. This is consistent with the view that participants were segmenting based on movement features for both fine-grained and coarse-grained viewings.

In short, Experiment 2 replicated the central finding from Experiment 1: Movement features were more predictive of fine-grained segmentation than coarse-grained segmentation. However, the data also suggested that participants in the second experiment may have relied more on movement features in general than the participants in Experiment 1. First, movement features accounted for more variance in segmentation overall. Second, the two groups showed equal agreement regarding the location of fine-grained and coarse-grained unit boundaries, and higher agreement overall. Third, participants rated the movie used in Experiment 2 as less goal-directed and more random than the movie in Experiment 1. The first two differences could be due in part to the larger sample size in Experiment 2: Larger samples lead to more reliable estimates of the breakpoint histograms, which should produce higher correlations overall. However, the data also suggested that the difference between the stimuli used in Experiments 1 and 2 might have had an impact. To generate a movie that was consistent with the "video game" cover story for the intentional condition, Experiment 2 used an animation that was

qualitatively faster and jerkier than that used in Experiment 1. One possibility is that sensory characteristics of the stimulus place important boundary conditions on observers' use of intentional attributions to segment activity. Laboratory personnel who viewed the stimuli from Experiment 1 and Experiment 2 reported an impression that the movement patterns in the Experiment 1 stimulus were more suggestive of intentional activity than those in the Experiment 2 stimulus. Consistent with this impression, participants in Experiment 2 rated the movie as less goal-directed and more random overall, compared to Experiment 1. However, this difference was not statistically reliable. ANOVAs comparing the two experiments directly found no main effect of experiment or condition by experiment interaction, (largest $F(1, 64) = 1.69$, $p = .20$).

Experiments 1 and 2 both provided support for the first two postulates of the model described in Section 1: Movement features were reliably related to the identification of fine event segments, and were less strongly related to coarse event segments. Post hoc, comparison of Experiments 1 and 2 suggests support for the third and fourth postulates. Participants' interpretations of the movies may have been affected by the differences between the stimuli (postulate 4), and this in turn may have affected how movement features were related to event segmentation (postulate 3). However, these effects could also have been due to incidental differences in the speed of the stimuli or in the cover stories used. To provide a stronger test, Experiment 3 manipulated the stimuli experimentally, comparing matched movies that either depicted goal-directed activity or random motion.

4. Experiment 3

In Experiments 1 and 2, participants were told that a movie depicted either goal-directed actions or random movement—but in both cases the movie was in fact randomly generated. The primary change in Experiment 3 was the manipulation of whether the movie actually depicted goal-directed human activity. To create movies of actual goal-directed activity, pairs of participants enacted several activities using the video game used in the cover story in Experiment 2. For each of the movies selected for the main experiment, a random motion movie was created that was matched for the speed and acceleration of the objects. Thus, there were two *stimulus* conditions: Participants in the *game* condition saw movies of other people playing the video game, and participants in the *equation* condition saw movies generated randomly using the equations used in Experiments 1 and 2.

The primary hypothesis was that participants would base their segmentation more on movement features during fine-grained segmentation, and more when told that the stimuli were random, as in the two previous experiments. A secondary hypothesis was that they would rely more on movement features when the stimuli were in fact randomly generated (the equation stimulus condition). As in the previous experiments, this was expected to lead to greater agreement between the intentional and random interpretation groups for fine-grained segmentation than coarse-grained segmentation. Finally, it was expected that there would be more agreement across interpretation groups for those who saw the equation stimuli than those who saw the game stimuli.

4.1. Method

4.1.1. Stimulus generation study

In order to generate movies for the game stimulus condition, sixteen participants, who were drawn from the same population as for the main experiments, were invited to participate in a pilot study, in partial fulfilment of a course requirement or for a $10/h payment. The video game used in Experiment 2 was modified slightly for this study. The two players each controlled either the orange circle or the green square and could move the objects around a 480-pixel white square. The keys to control the movement were as in Experiment 2. However, for the stimulus generation study two separate monitors and keyboards were attached to the computer, facing each other across a table. Each participant sat at one of the monitors and used their own keyboard. The experimenter demonstrated how the controls worked and explained that the purpose of the study was to capture representative examples of people using the video game to perform goal-directed actions. After this training, participants were asked to perform some of the intentional activities studied by Blythe et al. (1999):

> *Chasing*: One player was instructed to move their object so as to intercept the other player as quickly and often as possible, and the other player was instructed to avoid being intercepted.
> *Courting*: One player was instructed to move their object so as to court the other player's object by interacting with it in any way that it might find interesting, exciting, or enticing; the other player was instructed to play the role of the one being courted.
> *Fighting*: Each player was instructed to move their object to attack the other object from behind while avoiding being attacked from behind.
> *Following*: One player was instructed to follow the other's object as closely as possible, and the other player was instructed to move so as to encourage this following.
> *Guarding*: One player was instructed to choose a location on the screen to guard from the other player, and the other player was instructed to try to sneak into the guarded location.
> *Playing*: Both participants were instructed to play with the other participant in any manner they preferred.

After instructions and initial practice, each pair of participants performed three or six of the activities, for 300 s each. During the operation of the game, the position of each object was measured at 10 frames/s, and movies were generated from these object trajectories.

Consultation with members of the laboratory group resulted in selection of four examplars that appeared to be typical and good examples of the intended activity. The four activities selected were: chase, court, fight, and play.

For each of the video game movies, a random movie was generated that was matched as closely as possible for mean speed and acceleration magnitude of the objects. For the game movies, mean speeds ranged from .12 units/s to .33 units/s, and mean acceleration magnitude ranged from .36 units/s^2 to .95 units/s^2. For each of the four random movies the A and D parameters of Eqs. (1a) and (1b) were adjusted to generate matched motion trajectories. All trajectories were within .029 units/s of the target mean speed, and within .084 units/s^2 of the target acceleration.

In short, the result of the stimulus preparation study was four movies drawn from participants playing a video game, and four movies that were matched for their low-level properties.

Multiple stimuli were used here, rather than one longer movie, to provide increased generalizability. In the previous experiments item effects were not a serious concern because the movies were randomly generated. However, in the present case it is possible that incidental features of one of the intentional activities could have unduly influenced the results if only one activity were studied. (The movies are available in the Cognitive Science on-line annex; see Movies 3–10 at http://www.cognitivesciencesociety.org/supplements/.)

4.1.2. Participants

Sixty-two Washington University undergraduates (20 male, age 18–26) participated in partial fulfillment of a course requirement or in return for $10 compensation. None of the participants had participated in a previous study of event segmentation in our laboratory.

4.1.3. Materials

Animations were presented as in Experiments 1 and 2, using the same software and hardware setup. The materials used for the intentional and random interpretation conditions were the same as in Experiment 2: Participants in the intentional interpretation condition were given an opportunity to play the video game, and participants in the random interpretation condition were shown the equations used to generate the stochastic stimuli. In this Experiment, rather than using one brief training movie and one longer target movie, four movies of equal length (300 s) were used. One of the movies was used for training and the other three were used as experimental stimuli, with order counterbalanced using a Latin square.

4.1.4. Procedure

The procedure closely followed that of Experiments 1 and 2. Each person was randomly assigned to one of the two interpretation conditions and one of the stimulus conditions, resulting in four groups: random interpretation/equation stimulus (15 participants), random interpretation/game stimulus (16 participants), intentional interpretation/equation stimulus (15 participants), or intentional interpretation/game stimulus (16 participants). After providing informed consent, each participant was given instructions that corresponded to their interpretation condition. They were then given the opportunity to segment the first movie for practice. As before, if their unit size was outside of the desired normative range, they were given feedback and asked to repeat the practice movie. Participants who were not within the normative range after three attempts were excused. As in Experiment 2, use of this restrictive criterion provided a precise estimate of fine and coarse segmentation behavior, but at the cost of excluding participants who did not reach the criterion. A total of 22 participants were replaced, 4 each in the two intentional interpretation conditions, 11 in the random interpretation/equation stimulus condition, and 3 in the random interpretation/game stimulus condition. It is not clear what caused this higher rejection rate; one possibility is difference in the experimenters administering the protocol. The variability of rejection rate across conditions approached but did not reach statistical reliability, $\chi^2(3) = 7.45$, $p = .06$. Three participants successfully completed the training but produced more coarse than fine breakpoints during the experiment; these three were also replaced. An additional four participants were replaced due to equipment failure. Three others were missing data from one or two viewings; their remaining data were retained.

Each participant completed the training procedure at either a coarse or fine grain (counterbalanced across participants), and then completed the three remaining movies. They then were trained on the other segmentation grain and segmented the three target movies again. Participants also completed a questionnaire similar to that used in the first two experiments. However, administration of the questionnaire was modified to better accommodate the fact that multiple stimuli were used. Participants were asked to rate the goal-directedness and randomness of each movie immediately after the first viewing of each movie. (Participants were not asked to describe the activity in the movies, so as to avoid influencing their second segmentation.) At the end, they were asked to describe in their own words how they decide where small boundaries were, how they decided where large boundaries were, and whether they had anything else of interest to tell us. After completing these final questions, participants were debriefed and excused.

4.2. Results

4.2.1. Unit size

The sizes of coarse and fine units for all four groups are shown in Table 6. As in Experiment 2, the training procedure was successful in creating consistency in participants' grain of segmentation across experimental conditions. Each participant's mean number of breakpoints identified across the four movies for fine and coarse segmentation was calculated. These means were submitted to an ANOVA with interpretation and stimulus type as between-participants variables and segmentation grain as a repeated measure. As expected, there was a large effect of grain, $F(1, 58) = 127.5$, $p < .001$. None of the other main effects or interactions approached statistical reliability, largest $F(1, 58) = 2.06$, $p = .16$.

4.2.2. Role of movement features

Movement features were calculated as before, and forward stepwise regressions were used to estimate which movement features predicted where observers segmented under each of the eight experimental conditions. The regression results are reported in Table 7. As predicted,

Table 6
Mean breakpoint counts and unit sizes as a function of grain of segmentation, interpretation condition, and stimulus type in Experiment 3

Stimulus type	Grain	Interpretation	Number of breakpoints	Unit length
Game	Fine	Random	24.7 (11.5)	15.1 s (11.4 s)
		Intentional	24.0 (8.36)	13.2 s (4.27 s)
Game	Coarse	Random	9.61 (4.12)	33.9 s (16.5 s)
		Intentional	10.1 (4.93)	33.7 s (17.7 s)
Equation	Fine	Random	28.2 (20.5)	13.8 s (6.94 s)
		Intentional	28.7 (11.5)	12.2 s (7.23 s)
Equation	Coarse	Random	9.38 (5.41)	37.4 s (21.2 s)
		Intentional	10.3 (5.29)	36.1 s (26.1 s)

Standard deviations in parentheses.

Table 7
Relationship between movement features and segmentation in Experiment 3

Stimulus type	Grain	Interpretation	R^2 (95% conf. int.)	Most predictive features*
Game	Fine	Random	.20[b,c] (.16–.24)	+Distance Minima, −Distance, +Acceleration Magnitude$_C$, +Relative Acceleration
		Intentional	.19[b,c] (.15–.23)	−Distance, +Relative Acceleration, +Acceleration Magnitude$_C$
Game	Coarse	Random	.06[e] (.04–.09)	+Position$_{S,X}$
		Intentional	.05[e] (.03–.08)	+Relative Acceleration
Equation	Fine	Random	.22[b] (.18–.26)	+Acceleration$_S$, +Relative Acceleration, −Distance, +Speed$_S$
		Intentional	.31[a] (.26–.35)	−Distance, +Acceleration Magnitude$_S$, +Relative Acceleration
Equation	Coarse	Random	.10[d] (.07–.14)	+Acceleration Magnitude$_S$, +Relative Acceleration
		Intentional	.16[c] (.12–.20)	−Distance, +Acceleration Magnitude$_C$, +Relative Acceleration

For each combination of grain of segmentation, interpretation condition, and stimulus type, the table gives the proportion of variance accounted for by movement features (R^2). R^2 statistics that share a superscripts do not differ by Fisher's z transformation. The rightmost column lists the movement features that accounted for most of the variance, with a sign (+ or −) to indicate whether the correlation between that variable and segmentation was positive or negative.

* Features are listed in the order they entered the stepwise regression, up to the last feature that accounted for at least 1% of incremental variance. Abbreviations in subscripts refer to one of the two objects ("C" for circle, or "S" for square) or to one of the two dimensions of movement ("*X*" or "*Y*").

the differences between the game and equation movies mirrored the main difference between Experiments 1 and 2: Movement features accounted for more variance in the equation movies than in the game movies. For both movie types, movement features accounted for more of fine-grained segmentation than coarse-grained segmentation, replicating the pattern in the previous experiments. As in the previous experiments, the movement features that were most predictive of segmentation were related to the distance between the objects and to their acceleration. The interpretation manipulation had relatively little effect on the strength of the relationship between movement features and segmentation, relative to the other manipulations.

4.2.3. Agreement across conditions

Agreement between the groups' segment boundaries was compared using breakpoint histograms, as was done for the two previous experiments. It is not meaningful to calculate agreement between those groups who saw the game movies and those who saw the equation movies. Therefore, separate analyses were conducted comparing agreement between the intentional interpretation subgroup and the random interpretation subgroup for each of the two stimulus sets. For the groups that saw the game movies, those in the intentional and random attribution conditions showed good agreement in the placement of their fine-grained segment boundaries ($r = .63$, $p < .001$) and reduced, though still highly reliable, agreement in the placement of their

Table 8
Mean ratings of goal-directedness and randomness as a function of stimulus type and interpretation in Experiment 3

Stimulus type	Interpretation	Goal directedness	Randomness
Game	Random	3.75 (1.06)	4.04 (1.15)
	Intentional	5.02 (.97)	3.60 (1.20)
Equation	Random	3.67 (1.38)	4.40 (1.43)
	Intentional	3.91 (.97)	4.11 (1.11)

Standard deviations in parentheses.

coarse-grained boundaries ($r = .30$, $p < .001$). The difference in correlations was statistically reliable, $z = 10.4$, $p < .001$. For the groups that saw the equation movies, the two attribution subgroups agreed well about fine segment boundaries ($r = .45$, $p < .001$), and slightly better about the location of coarse boundaries ($r = .51$, $p < .001$). This difference, though smaller than that for the game movie groups, was also statistically reliable, $z = 1.80$, $p = .04$. In short, for the game movies the two interpretation groups agreed better about fine-grained boundaries than coarse-grained boundaries, whereas for the equation movies the two interpretation groups agreed slightly better about coarse-grained boundaries.

4.2.4. Questionnaire response

Results of the questionnaire Likert scales were analyzed as for the previous experiments. Mean ratings of goal-directedness and randomness were calculated for each participant and submitted to analyses of variance. The data are summarized in Table 8. As can be seen in the table, the game movies were rated more goal-directed and less random than the equation movies. Participants in the intentional interpretation group rated the stimuli as more goal-directed and less random than those in the random interpretation group. For the ratings of goal-directedness both the main effect of stimulus type and the main effect of interpretation were reliable, $F(1, 58) = 5.5$, $p = .04$, and $F(1,58) = 7.6$, $p = .008$, respectively. As the table indicates, the interpretation manipulation had a larger effect for those who saw the game stimuli than those who saw the equation stimuli; however, the interaction of interpretation and stimulus type was not statistically reliable, $F(1, 58) = 3.34$, $p = .07$. For the ratings of randomness none of the effects were statistically reliable, largest $F(1, 58) = 1.92$, $p = .17$.

4.3. Discussion

The results of this experiment supported the model shown in Fig. 1. First, movement features were reliably correlated with perceptual segmentation in all conditions. Second, these correlations were larger for fine-grained segmentation than for coarse-grained segmentation. Third, correlations were affected by manipulations designed to affect the knowledge structures observers brought to bear: Both the actual stimulus and the cover story affected the relationship between movement features and segmentation. Finally, both the stimulus and the cover story affected observers' judgments of the intentional structure of the activity.

To a first approximation, the results of Experiment 3 recapitulate those of Experiments 1 and 2. Movement features were more predictive of fine-grained segmentation than coarse-grained

segmentation. Movement features were most predictive of event segmentation for the movies that were perceived as less goal-directed (the equation movies) and less predictive for the movies seen as more goal-directed (the game movies). The direct manipulation of the stimulus characteristics strengthens the implications of these effects.

However, one aspect of the data stands out from the previous results. For the equation movies, there was a stronger relationship between event segmentation and movement features for those participants who were told that the activity was intentional than for those who were told the activity was random. This argues against the reasonable intuition that the relationship between movement features and intentions is "either/or," i.e., that as people base their segmentation more on inferred intentions they base their segmentation less on movement features. Rather, it suggests that inferring intentions may affect *how* movement features are processed. The particular pattern of which features were most predictive of event segmentation is consistent with this interpretation (see Table 7). For those who viewed the equation movies in the random interpretation condition, the square's acceleration was the best single predictor for both fine and coarse segmentation ($r = .31$ and .16, respectively), whereas the distance between the objects was less predictive ($r = -.15$ and $-.14$, respectively). However, for those in the intentional interpretation, the square's acceleration was less predictive ($r = .19$ for fine, $r = .11$ for coarse) and the distance between the objects was more predictive ($r = -.41$ for fine, $r = -.31$ for coarse). (For both of these movement features, the difference in correlation between the random and intentional groups was tested using the method for comparing correlated correlation coefficients described by Meng, Rosenthal, & Rubin (1992). This was performed separately for fine and coarse segmentation. All four pair-wise differences were statistically reliable, $p \leq .047$.) One possibility is that participants in the random interpretation condition attended to the individual movements of one of the objects at a time, whereas participants in the intentional interpretation condition attended to the relationship between the objects. This result indicates that for those who saw the equation movies, how movement features related to event segmentation depended on how they interpreted the movies.

5. Temporal aspects of movement-perception coupling

The analyses reported in the previous sections speak to the strength of the relationship between movement features and event segmentation, and to the particular aspects of movement that are related to segmentation. Those analyses were performed to test the predictions of the model described in Section 1. However, the data also speak to the temporal relationship between movement information and perceptual segmentation, which has not been previously characterized. This section briefly summarizes the results of analyses characterizing the temporal relationship between movement features and segmentation.

In all three experiments, the temporal relationship between each movement feature and perceptual segmentation was calculated. As noted previously (see Section 2.2), for each stimulus set the cross-correlation between probability of segmentation and each movement feature was calculated. (Cross-correlations were computed after collapsing the segmentation data across grain and interpretation. This was done by dividing the breakpoint histogram for each grain and interpretation condition by its standard deviation, to control for differences in the num-

ber of breakpoints reported, and then summing the weighted breakpoint histograms.) Lags from −5 s to 5 s were examined. Positive lags correspond to reactive coupling, such that a movement feature's value at a given time correlates best with segmentation at a later time. Negative lags correspond to anticipatory coupling, such that a movement feature's value at a given time correlates best with segmentation at an *earlier* time. For each movement feature for each stimulus set, the optimal lag and its corresponding correlation was recorded. Only lag values corresponding to variables that were related to perceptual segmentation are of interest; for non-predictive features, the lag values are essentially random. Therefore, to characterize the temporal coupling between predictive movement features and segmentation, the lags for each of the movement features listed in Tables 3, 5, and 7 were tabulated. (As in those tables, each feature was counted each time it entered the regression equation for a given condition, accounting for more than 1% of incremental variance.)

For the 57 features that accounted for incremental variance in event segmentation across the three experiments, 35 had lags of zero (see Fig. 3). That is, there was generally a very close temporal coupling between a movement feature and its effect on perceptual segmentation. However, as can be seen in the figure, non-zero lags were also observed. Most of these were positive, reflecting that the effect of a movement feature on observers' segmentation was strongest 1–5 s later. However, 7 of the recorded lags were negative, indicating anticipatory coupling. There were no obvious relationships between the different types of movement features and the nature of the lag observed.

At first blush, such anticipatory coupling would appear impossible, an instance of backward causality in time. However, the dynamics of the movement features in both the equation and game stimuli included substantial temporal autocorrelation. In domains with autocorrelation,

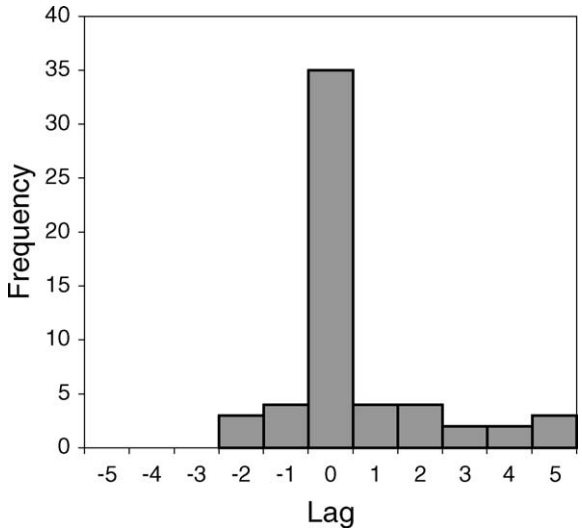

Fig. 3. The temporal relationship between movement features and their effect on event segmentation. The histogram plots the frequency of lags between movement features and perceptual segmentation. Those features that accounted for more than 1% of incremental variance in perceptual segmentation in each condition across the three experiments are included in the frequency counts.

anticipatory coupling is not only possible, but common (see, e.g., Kelso, 1995). Thus, one possibility is that viewers anticipate the effects of a change in a movement feature based on what is likely to follow that change. Another possibility is that anticipatory lags reflect the performance of participants during their second viewing of a stimulus, during which they might segment based on memory for what is about to happen. Of course, it is also possible that the small number of negative lags observed here simply reflect random variation. Characterizing the reliability of the presence of anticipatory coupling, and the circumstances under which it occurs, is an important problem for future research.

6. General discussion

The three experiments reported here independently manipulated physical characteristics of movement and information about goals and intentions, in order to evaluate their effects on viewers' perceptual segmentation. The results provide clear answers to the two questions raised in Section 1. First, to what degree do movement features determine how viewers perceive event parts? Under some circumstances, movement features appear to be strongly related to the perception of event parts. For fine-grained segmentation of animations that appeared random (Experiment 2 and the equation movies of Experiment 3), movement features accounted for as much as 33% of the variance in where observers segment activity. Limits on the amount of variability in segmentation that can be accounted for come from two sources. First, there are stable individual differences in segmentation (Speer et al., 2003), which in this case may reflect in part differences in participants' interpretation of the objects' intentions. Second, any individual segmentation of an activity includes noise that may reflect momentary lapses of attention, hesitations, or variability in the latency to press the button. Given these limits, R^2 statistics above .3 can be considered strong relationships.

However, under other circumstances, movement features were only weakly related to event boundaries. For coarse-grained segmentation of animations that appeared less random (Experiment 1 and the game movies of Experiment 3), movement features for accounted substantially less of the variance in where observers segmented the activity. This result is more surprising because there is an important sense in which the viewers had nothing to base their segmentation on other than movement. The stimuli had none of the other typical cues to event structure: language, facial expression, props, or setting. Previous results suggest that interactions with objects are particularly important for coarse-grained segmentation (Zacks, Tversky, et al., 2001), but in these simple stimuli there were no objects with which to interact. Even in this reduced setting it appears that viewers can abstract from the stimulus such that their segmentation is not predicted by the exhaustive coding of sensory characteristics used here.

Perhaps viewers simply were more idiosyncratic when identifying coarse-grained event boundaries, and this explains why movement features were less effective for predicting coarse-grained segmentation? This clearly cannot explain the difference, because the reduced predictiveness of movement features for coarse grained segmentation held both when cross-group agreement was higher for fine segmentation (Experiment 1, Experiment 3 game movies) and when it was higher for coarse segmentation (Experiment 3 equation movies).

Now, to the second question raised previously: Does knowledge of actors' intentions modulate the role movement features play in segmenting events? This question was addressed by two experimental manipulations. First, participants in all three experiments were told either that the activity was randomly generated or goal-directed. This significantly affected the relationship between movement features and event segmentation for the equation movies in Experiment 3. Second, the degree to which the stimuli shown gave the appearance of goal-directed activity varied. This variation was haphazard in comparing Experiments 1 and 2, and experimentally rigorous in comparing the game and equation movies of Experiment 3. In both comparisons, stimuli that gave more cues to being intentional activity produced weaker relationships between movement features and event segmentation.

The most robust finding in these experiments is that movement features predicted fine-grained segmentation better than coarse-grained segmentation. This result is consistent with the previously described observations by Newtson et al. (1977), using different measures of movement and realistic movies. By itself, this suggests a simple and intuitive picture in which bottom-up processing of sensory characteristics determines fine units, but top-down processing based on knowledge structures determines coarse units. On this account, the two sources of influence would be antagonistic, such that as the influence of movement features on segmentation increased, the influence of intentions would wane. Recall, however, that knowledge structures influence perception by modulating the processing of sensory characteristics. (Else, one is dealing with memory or imagination rather than perception.) This argues against the simple antagonism view, and suggests that although effects of intentions on movement features may decrease as the influence of movement features increases, they also could increase, or change their form without increasing or decreasing. Increases, decreases, and changes in form were all observed here.

After the differences between coarse and fine units, the largest effects on the relationship between movement features and segmentation were produced by stimulus changes. In Experiment 3 in particular, closely matched random and game movies produced quite different relationships between movement features and segmentation. Manipulating the interpretation provided to viewers had much less of an effect on their segmentation. However, ratings of the goal-directedness of the stimulus showed the opposite pattern: They were more affected by viewers' interpretations than by changes to the stimulus. One possibility is that this reflects a dissociation between viewers' reflective, explicit evaluation of the degree of intentional structure in the movie and viewers' moment-to-moment, implicit reactions. Another possibility is that this simply resulted from a set of demand characteristics that encouraged participants to rate the allegedly human-generated movies as more goal-directed. The possibility that implicit measures such as event segmentation may tap into viewers' perceptions of intentions merits further investigation.

Thus, the data permit two broad conclusions about the role of movement features in event perception. First, under some (but not all) circumstances, movement features can play a major role in event segmentation. Second, the role played by movement features is modulated in a top-down fashion by inferences about actors' intentions. More detailed conclusions come from the model presented in Section 1. This model makes four postulates, which were supported by the data: First, movement features contributed to the identification of fine event segments. This was consistently true across the three experiments. Second, in all three experiments observers appeared to rely less on movement features as the grain of segmentation increased. Third,

inferences about actors' intentions modulated the relationship between movement features and event segmentation. Finally, inferences about actors' intentions were affected both by intrinsic features of the stimuli (differences between the movies), and by top-down information provided by the interpretation manipulation. The model supports a view of event perception in which the bottom-up processing of sensory characteristics is modulated by top-down control processes.

Across the three experiments, the temporal coupling between a change in a movement feature and its effect on segmentation was often nearly instantaneous. However, in a minority of cases the effect of a movement change was delayed in time. There was also evidence for anticipatory effects, such that participants segmented in advance of a coming change in a movement feature, but these effects could be artifactual and should be investigated further.

Previous research on the basis of event segmentation has studied sensory characteristics or knowledge structures in isolation. The present results indicate that, in order to achieve a complete understanding of event perception, researchers cannot ignore interactions between the two. This is a knotty problem, because sensory characteristics such as movement features tend to covary with aspects of knowledge structures such as goals. In the present studies these were teased apart somewhat by manipulating the cover story provided to viewers, and by presenting simple stimuli closely matched for overall movement information. This approach opens up other strategies for research. For example, one could construct movies that are ambiguous with regard to their goal structure, and provide different participants with different interpretations, allowing for finer-grained assays of the effects of intentional structure on event segmentation.

The covariation of sensory characteristics of activity and knowledge structures for representing events is more than a methodological bugbear. It points to an important conclusion about human perception of dynamic stimuli: People are adapted to pick out those aspects of the physical structure of ongoing activity that provide reliable information for achieving conceptual understanding.

Notes

1. The participants also tended to segment the movies at those points where successive shots were temporally discontinuous. However, it is not clear how to generalize this to event perception, because temporal discontinuities are an artifact of film editing.
2. This resulted from the retention of the ∂ values in the no-collision and wall-avoiding constraints. When either object encountered an obstacle, both objects had a tendency to change direction on the next frame, because the ∂ parameters had drifted farther than on no-collision time-steps. This was true for the randomly generated animations in Experiments 2 and 3 as well.

Acknowledgments

This research was supported in part by the James S. McDonnell Foundation. The experiments were conducted with invaluable assistance from Mary Dowd, Bart Phillips, Margaret

Sheridan, and Jean Vettel. Pascal Boyer, Pascale Michelon, Nicole Speer, Khena Swallow, Barbara Tversky, and Jean Vettel provided thoughtful feedback during the writing of the manuscript. Thanks to Peter Todd for providing stimuli and suggestions for the video game paradigm.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cogsci.2004.06.003.

References

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

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. New York: The Macmillan Company.

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology, 33*, 680–685.

Blythe, P. W., Todd, P. M., Miller, G. F., et al. (1999). How motion reveals intention: Categorizing social interactions. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 257–285). New York, NY, USA: Oxford University Press.

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*, 257–271.

Friston, K. J., & Buchel, C. (2000). Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 7591–7596.

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

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18*, 65–96.

Kelso, J. A. S. (1995). *Dynamic patterns: Self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*, 533–545.

Meng, X. -l., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175.

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

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: L. Erlbaum Associates.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). On the role of human areas MT+ and FEF in event perception. *Cognitive, Affective and Behavioral Neuroscience, 3*, 335–345.

Wilder, D. A. (1978a). Effect or predictability on units of perception and attribution. *Personality and Social Psychology Bulletin, 4*, 281–284.

Wilder, D. A. (1978b). Predictability of behaviors, goals, and unit of perception. *Personality and Social Psychology Bulletin, 4*, 604–607.

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

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.