

Making sense of abstract events: Building event schemas

BRIDGETTE M. HARD, BARBARA TVERSKY, and DAVID S. LANG
Stanford University, Stanford, California

Everyday events, such as making a bed, can be segmented hierarchically, with the coarse level characterized by changes in the actor's goals and the fine level by subgoals (Zacks, Tversky, & Iyer, 2001). Does hierarchical event perception depend on knowledge of actors' intentions? This question was addressed by asking participants to segment films of abstract, schematic events. Films were novel or familiarized, viewed forward or backward, and simultaneously described or not. The participants interpreted familiar films as more intentional than novel films and forward films as more intentional than backward films. Regardless of experience and film direction, however, the participants identified similar event boundaries and organized them hierarchically. An analysis of the movements in each frame revealed that event segments corresponded to bursts of change in movement features, with greater bursts for coarse than for fine units. Perceiving event structure appears to enable event schemas, rather than resulting from them.

When people view a bustling city street, they perceive individual objects: cars, stoplights, and other people. When those objects move, people perceive individual *events*. Events come in many varieties; they can involve animate and inanimate entities, short and long periods of time. But much of what we perceive in everyday life are behavioral events, or goal-directed activities involving people: a person buying a hot dog from a street vendor, or a driver turning at an intersection. People thus make sense of their experiences by extracting meaning from a continuous flow of information.

Despite the inherently dynamic nature of events, perception of objects and events share many characteristics. Like objects, events are selected and segmented from ongoing information. In other words, they are perceived as discrete, as individually distinct and *bounded*, with beginnings and ends (e.g., Casati & Varzi, 1996; Newtonson & Engquist, 1976; Zacks, Tversky, & Iyer, 2001). Like objects, events are also conceived as having hierarchical structure, as being composed of parts, or subevents. Buying a hot dog, for example, consists of placing an order, paying the vendor, and receiving the tasty treat. Relative to the perception of objects, however, the perception of

events and their structure is not well understood. How is it accomplished?

One approach to this question has been to examine people's explicit knowledge, or *schemas*, for common events. This approach has confirmed that when people list event constituents, they organize them in partonomic, or whole-part, hierarchies (e.g., Abbott, Black, & Smith, 1985; Bower, Black, & Turner, 1979). These hierarchical organizations generally correspond to the goals and subgoals of the event, as in the example of buying a hot dog.

A complementary approach has been to study how event structure is detected online. In this approach, participant observers are asked to segment the living stream of behavior (Barker, 1963) or films of it (Newtonson, 1973) into natural units. Although there is inevitable variability in how observers segment behavior into events, they are remarkably consistent; they agree with each other on the locations of unit boundaries, called *breakpoints* (Dickman, 1963; Newtonson & Engquist, 1976; Zacks et al., 2001).

Observers segmenting ongoing events also organize them hierarchically, consistent with explicit event knowledge. In one set of studies that demonstrated this, observers viewed and segmented everyday events, such as making a bed or doing the dishes, into large and small units (Zacks et al., 2001). Some observers gave a verbal play-by-play as they segmented, allowing analysis of their explicit interpretations of the events. Analysis of segmentation patterns showed that the breakpoints of larger units aligned with those of smaller units. This *hierarchical alignment effect* suggested that the events were perceived as partonomic hierarchies. Descriptions of the events were consistent with this interpretation; the observers described events involving completion of goals and subgoals.

Why do people organize their understanding of behavioral events according to partonomic hierarchies? One

This research was supported by Office of Naval Research Grants N00014-PP-1-O649, N000140110717, and N000140210534 to Stanford University. We thank Helen Hwang for her help with data collection. We also thank Ewart Thomas, Gordon Bower, Herb Clark, Jeff Zacks, Dan Yarlett, Rolf Zwaan, and an anonymous reviewer for their insightful comments and suggestions. Correspondence concerning this article should be addressed to B. M. Hard, Department of Psychology, Stanford University, Jordan Hall (Bldg. 01-420), Stanford, CA 94305 (e-mail: hard@psych.stanford.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

reason is that behavior is, in fact, produced hierarchically. In a classic paper, Lashley (1951) argued that hierarchical organization is fundamental for all skilled acts, down to the level of coordinating muscle contractions for reaching and grasping; constituent acts simply occur too quickly to allow sequential chaining. The everyday event of making a bed offers one example of a hierarchically organized behavior. Individual but functionally related movements combine to form small actions such as lifting the sheet. Small actions are combined with others to achieve larger goals. For example, lifting the sheet, spreading it out, tucking in the corner, and so forth combine to achieve a single goal: put the sheet on the bed. At an even larger scale, actions such as putting the sheet on the bed combine with other actions, such as putting on the blanket, to achieve the larger goal of making the bed or, on an even larger scale, cleaning the house. Hierarchical organization characterizes action coordination in problem solving and planning, as well as in the performance of routine, everyday tasks. In *means-ends analysis*, or *operator subgoaling*, people solve problems by establishing goals and then partitioning those goals into component subgoals (Newell & Simon, 1972). In sum, the fact that people perceive observed activities hierarchically suggests that they infer or detect the inherent hierarchical structure of others' actions.

The idea that observers *infer* the hierarchical structure in others' actions suggests one interpretation of Zacks et al.'s (2001) findings: that online event perception is directed top down by explicit knowledge about goal hierarchies. In other words, observers segment observed activities according to preexisting event schemas that encode goals and subgoals. Consistent with this hypothesis, Zacks et al. showed that when observers were more familiar with an event sequence, hierarchical alignment increased. Hierarchical alignment also increased when the observers described while segmenting. Because describing required the participants to justify their segmentation choices, it probably encouraged them to better interpret each event using explicit knowledge of the activity, leading to more organized segmentation overall.

But does the online perception of hierarchical structure depend on event schemas, or can structure be *detected* bottom up from objective information in observed activity? This question has two parts. First, how much do people rely on event schemas to segment experience into discrete events? Second, how much do people rely on event schemas to organize these segments hierarchically?

Event schemas clearly play a role in how experience is segmented into discrete events, by enabling people to anticipate certain events and not others (Neisser, 1976). But with or without event schemas, bottom-up sensory information offers objective cues to event boundaries (Newton, Engquist, & Bois, 1977). Specifically, natural discontinuities in the behavior stream correlate with and, thus, cue actors' intentions (e.g., Zacks & Tversky, 2001; Zacks et al., 2001). This analysis has parallels in object perception and cognition. Objects are naturally segmented into parts by salient breaks in contour (e.g., Biederman, 1987; Hoffman & Richards, 1984). Different parts serve different functions, so that perceptual discontinuities dis-

tinguishing parts correlate with and cue differences in function (Tversky & Hemenway, 1984).

What do discontinuities in the behavior stream look like? According to Barker and Wright (1955), discontinuities in behavior correspond to pauses as the actor completes one goal and shifts to another. Supporting the role of pauses in event segmentation, infants can individuate and even enumerate sequences of actions (i.e., jump-fall-jump) when those actions are separated by brief pauses but have more trouble when the actions form a continuous stream of motion (Sharon & Wynn, 1998; Wynn, 1996; but see Clearfield, 2004). Preschoolers are also drawn to pauses as a basis for segmentation, even when they are asked to ignore pauses in favor of more conceptual cues that correspond to goal completion (Wagner & Carey, 2003).

According to Newton et al. (1977), event boundaries are better characterized as bursts in movement complexity, rather than as pauses. Newton et al. asked naive observers to watch filmed sequences and divide them into natural units. For each film, 17 different features of the actor's body position were then coded, using dance notation. Event breakpoints corresponded to moments when the actor's body underwent relatively sudden changes in position. In everyday activities, both pauses and bursts in movement complexity may offer relevant cues to event boundaries. Furthermore, they are probably related. For example, when people pause to plan an action, their body position probably changes dramatically as they come to rest and then reorient toward the next goal.

Physical changes cue event boundaries, but do they also cue *hierarchical* event structure? Researchers in developmental (Baldwin & Baird, 1999) and comparative (Byrne, 2002) psychology have suggested that bottom-up information is sufficient for perceiving hierarchical organization in behavior. For example, infants who lack adult-level intentional knowledge nevertheless perceive certain combinations of movements, such as reaching for and grasping an object, as a unified act (Woodward, 1998). Infants can also segment continuous behavior according to points of major goal completion (Baldwin, Baird, Saylor, & Clark, 2001). Furthermore, rats, gorillas, and chimpanzees, who lack explicit knowledge of goal hierarchies, can imitate complex hierarchical organization in observed behavior (Byrne, 1999).

A physical basis to hierarchical event perception might be useful or necessary in building new event schemas. People, especially children, observe events that are novel, for which they have no event schemas and in which intentions and goals are difficult to infer. In these situations, people might exploit the coupling of physical and conceptual breaks to determine appropriate perceptual units. Different quantities or qualities of changes might correspond to different levels within an event hierarchy: goals and subgoals. Once perceptual units are identified, they are used to build event schemas. If this formulation is correct, observers should be able to segment events hierarchically even when they have no event schemas to guide them.

Alternatively, perceptual breaks in behavior might cue event boundaries, but not hierarchical structure. In support

of this proposal, segmentation becomes finer when events are less familiar, more ambiguous, or less predictable (Newson, 1973; Vallacher & Wegner, 1987; Wilder, 1978a, 1978b), suggesting that observers need knowledge about goals and intentions to group smaller units into larger ones. According to this formulation, events should be perceived more hierarchically when observers interpret the function or goals underlying sequences of events. This analysis is consistent with the findings of Zacks et al. (2001).

The aim of the present study was to test these alternatives by manipulating event knowledge and examining its effects on hierarchical segmentation. To accomplish this, we first considered that previous work, for the most part, had focused on familiar, concrete events enacted by a single person, typically involving actions on objects. Adults have extensive experience enacting and observing such events, making it difficult, if not impossible, to isolate the role of event schemas in segmenting them.

An appealing alternative to such events is the classic Heider–Simmel (1944) film. In that study, observers watched an animated film of one large and two small geometric figures behaving according to a “chase” scenario. Many people interpreted the movements of the geometric figures as humanlike, intentional actions (Heider & Simmel, 1944; Oatley & Yuill, 1985). They described similar stories involving the large triangle’s bullying the two smaller shapes: chasing them and trying to capture them. Thus, this animation can be interpreted according to a familiar, hierarchically organized event schema. But the events in this animation are abstract, in the sense that they lack many of the rich cues that accompany everyday events and constrain their interpretation. Instead of people, the agents are geometric figures; they have no limbs that make articulated movements or faces that turn and make expressions. Acts such as bullying and hiding are accomplished without limbs, heads, postures, or facial expressions. The scene in which these events unfold is also abstract, not a recognizable environment such as a kitchen or a park. Environments or scenes are excellent cues for the events that typically take place in them (Tversky & Hemenway, 1984). Removing these many rich cues renders the events more ambiguous. In comparison with natural events, constructing interpretations of these abstract events likely requires more effort and more top-down information based on prior experience. This makes understanding of abstract events open to manipulation.

Taking advantage of this fact, the present research manipulated the interpretability of abstract events to address several questions. What respective roles do event schemas and physical cues play in hierarchical event perception? Can events be perceived hierarchically even when there are no event schemas to guide understanding? If so, is the degree of hierarchical perception affected by whether observers interpret events as goal–subgoal hierarchies?

To answer these questions, we extended the methods and analysis used by Zacks et al. (2001). Observers segmented two films, one based on Heider and Simmel’s (1944) “chase” involving geometric figures, and one based on “hide-and-peek.” We manipulated the interpretability of

the events in two ways. First, the films were shown forward to some observers and in reverse to others. If events are bounded by attainments of goals, which are unidirectional in time, viewing the films backward should make goal attainment and, thus, goal hierarchies difficult to identify. Film direction should thus influence whether people can rely on familiar schemas to interpret the events. Second, we manipulated the observers’ familiarity with the events through repeated viewings and instructions to write a narrative prior to segmentation. This gave the observers an opportunity to build a specific schema for these events before segmenting them; inventing a story induced the viewer to think globally about the relationships among the events, which should facilitate later inferences about the functions or goals that underlie them.

As in Zacks et al.’s (2001) studies, half the observers described as they segmented. The primary aim of this manipulation was to collect descriptions so that the observers’ explicit interpretations of the events could be assessed. A second aim in collecting descriptions was to determine the effects of describing on hierarchical segmentation. Describing has previously been shown to improve hierarchical segmentation (Hard, Lozano, & Tversky, 2006; Zacks et al., 2001). Requiring observers to justify their segmentation with descriptions likely encourages them to make better sense of each event as it is happening, perhaps by relying more on prior event schemas that also guide the overall structure of the segmentation.

If hierarchical segmentation is influenced by event schemas, forward films should induce better hierarchical alignment than will backward films, and observers viewing the events for the first time should segment less hierarchically than will observers who are more familiar with the events. To preview the outcome, we found that although interpretations changed with event direction and familiarity, hierarchical segmentation did not. In fact, the observers selected remarkably similar event boundaries despite dramatic differences in interpretations. As a result of these findings, we examined the physical changes that predicted event boundaries and found evidence for an objective physical means of detecting event structure.

METHOD

Participants

Eighty Stanford undergraduates completed the experiment in exchange for course credit.

Films

Two 84-sec animated films were created with the animation program CuriousLabs Poser 4 (see Figure 1). Each animation was 2-D and portrayed three shapes, or “characters,” interacting with one another and with geometric “landmarks” in the environment. One animation, *chase*, was based on the script of the Heider and Simmel (1944) film, in which a large shape appears to bully and chase two smaller shapes. In the second animation, shapes appeared to play *hide-and-peek*. There was an original version and a perfectly reversed version of each film.

Design

The design was a $2 \times 2 \times 2$ mixed factorial, as summarized in Table 1. Film direction (*forward* or *backward*) was varied within

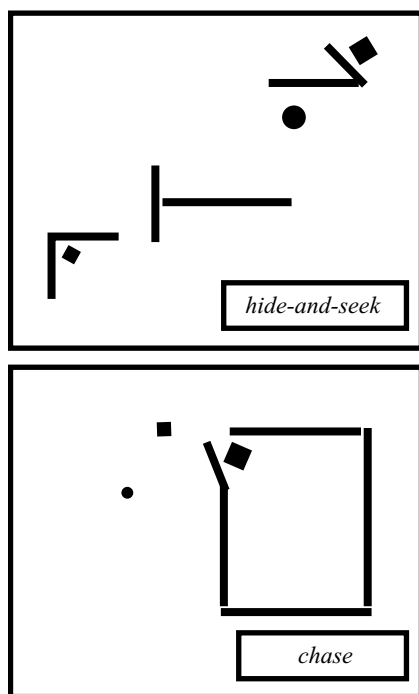


Figure 1. Still frames from the two stimulus animations: *hide-and-seek* (top) and *chase* (bottom).

participants. Familiarity (*novel* or *familiar*) and language during segmentation (*describe* or *silent*) were varied between participants. Each participant viewed the two films (*chase* and *hide-and-seek*), one forward and one backward, and segmented each film twice, once into coarse and once into fine units. The direction of each film, presentation order, and order of segmentation (fine/coarse or coarse/fine) were counterbalanced across conditions. The observers in the novel condition ($n = 48$) had no experience with the films prior to segmenting them. The observers in the familiar condition ($n = 32$) previewed the films multiple times and wrote brief descriptions of them before performing the segmentation task.

Procedure

The participants in the novel condition were first given instructions for segmenting the films. The participants in the familiar condition first viewed each film five times consecutively (in the order in which the films would later be segmented). They were told to pay

careful attention to the films, because they would later be asked to describe what happened in them. They were not told how to interpret them, nor were they told that they would later segment them. After five viewings, the participants were asked to write a brief description of each film. After writing their descriptions, the participants in the familiar condition were given instructions for segmenting the films.

For segmentation, the participants were instructed that they would see several short cartoons portraying geometric figures in motion. They were told to press the space bar whenever, in their judgment, one meaningful event ended and a new one began; the film would play continuously while they performed this task. The participants in the describe condition were also told that each time they pressed the space bar, they should briefly describe what happened in the segment they had just observed. On the first viewing, half of the participants were instructed to segment films into the smallest units that seemed natural and meaningful to them (fine units). The other participants were instructed to mark off films into the largest units that seemed natural and meaningful (coarse units).

The participants then viewed and segmented an 84-sec practice film depicting a game of *freeze tag* (the participants were not told that this was what the film depicted). The participants then had the opportunity to ask questions and continued with the rest of the experiment. Films were presented on a 21-in. flat screen computer monitor.

After segmenting the films, the participants completed an unrelated task for 10 min. They then segmented the two films a second time in the same order, using the opposite unit-size instructions. Response times were recorded on a Macintosh G4 computer attached to a keyboard, using a program written in PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). Verbal responses for the describe group were recorded using a tape recorder.

Two participants were excluded from analysis for indicating very large units that fell well outside the distribution for the group. In both cases, the participants were in the novel condition and indicated only one coarse unit for the backward film. The data of the remaining 78 participants were analyzed for segmentation and description.

RESULTS

Event Interpretations

The aim of this study was to separate top-down and bottom-up influences on hierarchical event perception. Film direction, simultaneous description, and familiarity were manipulated to assess their effects on segmentation, hierarchical organization, and interpretation. Top-down event schemas organize events in terms of goals and sub-goals. Participants using top-down interpretations, then,

Table 1
Experimental Design: $2 \times 2 \times 2$ Mixed Factorial,
With Familiarity and Language as Between-Subjects Factors
and Film Direction As a Within-Subjects Factor

Familiarity	Language	Forward Film	Backward Film
Novel (view once)	Silent	Chase	Hide-and-seek
	Silent	Hide-and-seek	Chase
	Describe	Chase	Hide-and-seek
	Describe	Hide-and-seek	Chase
Familiar (view five times)	Silent	Chase	Hide-and-seek
	Silent	Hide-and-seek	Chase
	Describe	Chase	Hide-and-seek
	Describe	Hide-and-seek	Chase

Note—Each film (*chase* and *hide-and-seek*) was shown forward for half the participants. The order of the films and the order in which the participants segmented into coarse and fine events were also counterbalanced.

should describe the events more intentionally, and less in terms of physical movements. Event schemas should be more readily invoked when events are familiar and when they are viewed in their natural forward direction. To preview the analyses of event descriptions given below, forward films were interpreted as more intentional than were backward films, and familiar films as more intentional, hierarchical, and complex than were novel films.

Event Categories

Descriptions were transcribed and divided into clauses, more or less corresponding to propositions. Clauses were then coded for intentional action, physical movement, and number of actors. Clauses were coded as *intentional* if the rater believed that they implied an intentional action, performed by a living being with goals and intentions. Clauses were coded as *nonintentional* if the rater believed that they could describe an action made by an inanimate object. For example, *hide*, *chase*, and *talk* were coded as intentional, whereas *move*, *rotate*, and *change direction* were coded as nonintentional. Clauses were coded as *physical movement* if the verb strongly implied motion—for example, *move*, *spin*, or *chase*. Given that these stimuli involved three geometric shapes whose movements are often contingent on one another, we were also interested in how often the participants described events involving single or multiple agents. Clauses were thus coded for the number of characters involved in the action (one or more than one).

Two coders classified 1,866 clauses from 44 participants. The descriptions of 4 participants could not be analyzed due to recording error. By Cohen's kappa, interrater agreement ranged from 84% to 97%, significantly above chance ($p < .01$) for all the categories. For each participant, the proportion of clauses in each category was determined for each film viewing (forward/backward and coarse/fine). For each category (intention, movement, and number of agents), these proportions were submitted to mixed factorial ANOVA with segmentation level, event direction, and familiarity as factors. The results are summarized in Figures 2 and 3 and will be described below.

Intentionality. If forward films are more interpretable in terms of goal-based event schemas than are backward films, forward films should be described as more intentional than backward films. Intentionality should be expressed especially in the verbs. As was predicted, the mean proportion of intentional descriptions was higher for forward films than for backward films [$F(1,32) = 5.10$, $p < .05$; see Figure 2]. If familiarity with the films enables better use of event schemas, familiar films should be described as more intentional than are novel films. As Figure 2 shows, increasing familiarity with the films increased the proportion of intentional descriptions [$F(1,32) = 13.96$, $p < .01$]. Also, coarse events were interpreted as more intentional than were fine events [$F(1,32) = 11.48$, $p < .01$; see Figure 3].

As can be seen in Table 2, descriptions in the novel condition overall were dominated by only a few verbs—specifically, *move* and *go*. In the familiar condition, verbs such as *move* and *go* were often replaced by intentional

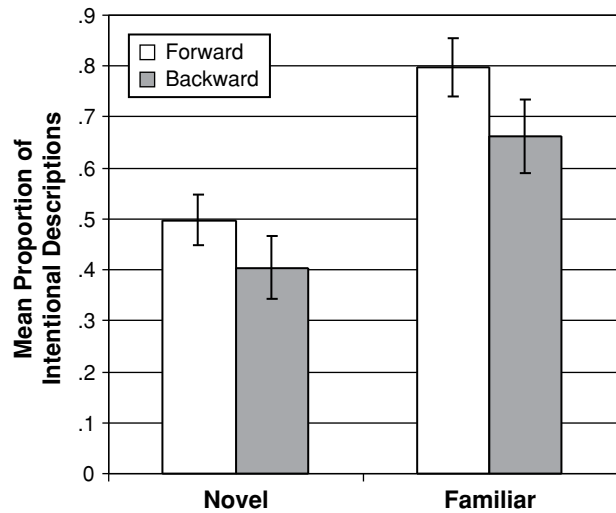


Figure 2. Mean proportion of intentional descriptions and standard errors for forward and backward films, viewed once (novel) or five times (familiar).

verbs, such as *run*, *leave*, and *enter*. Verbs that implied mental actions (*look* and *try*) also became more frequent. Table 2 shows only verbs that occurred in at least 1% of all the utterances; the participants in the familiar condition used a variety of other intentional verbs that are not depicted here because different participants chose subtly different verbs. Use of intentional and mental action verbs shows that the describers in the familiar condition were interpreting the events as causal and deliberate. The participants in the familiar condition also used the stative verb *be* more often than did the participants in the novel condition. The participants almost always used the stative verb in one of three ways: to set the scene by specifying a location (“the larger square is now inside the box”), to describe accomplishments (“that’s the end of the game”), or to specify a mental state (“the square is rather confused”). Participants in the familiar condition were thus better able than the participants in the novel condition to comprehend the setting, to interpret the events relative to the sequence as a whole, and to describe the mental states of the agents.

On average, a little more than half of the participants’ descriptions were of intentional actions ($M = .59$). This contrasts sharply with interpretations of familiar events such as doing the dishes, where fully 96% of the segments were described as intentional (Zacks et al., 2001). This difference is probably due to the more abstract nature of these events and the fact that these agents were geometric shapes, instead of people.

Movements. Because these stimuli involved simple geometric shapes and sparse scenes, we expected that the participants would focus primarily on physical movement in interpreting and segmenting the events, rather than on the properties of the characters or of the scene, and that this would be reflected in their descriptions. As was predicted, the majority of the descriptions reported

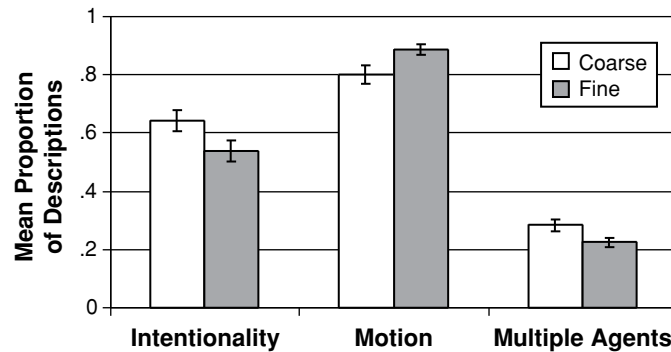


Figure 3. Mean proportion of descriptions and standard errors indicating intention, motion, and multiple actors for coarse and fine segmentation levels.

physical movement ($M = .85$). Film direction did not affect the mean proportion of movement descriptions, suggesting that the participants in both conditions focused equally on physical movement in interpreting the events [$F(1,32) = 1.53, p = .23$]. Film direction interacted with segmentation level, however [$F(1,32) = 4.52, p < .05$]. At the coarse level, the backward films elicited higher proportions of movement descriptions than did the forward films (.84 to .77, $SEM = .04$). This result indicates that the participants interpreted the backward films as less intentional and more physical than the forward films. At the fine level, however, the forward and the backward conditions did not differ (.89 to .89, $SEM = .02$), probably because fine-level descriptions were so predominantly movement based. As is evident from Figure 3, fine segmentation instructions led to more motion verbs than did coarse segmentation instructions [$F(1,32) = 13.93, p < .001$]. Together with more intentional descriptions for coarse units, this result suggests that fine-grained seg-

ments are understood more in terms of physical motion than are coarse-grained segments.

Finally, the participants who were more familiar with the films described less movement than did the participants who segmented on their first viewing [$F(1,32) = 6.95, p < .05$]. Increased experience thus led to more intentional and fewer physical interpretations of the events.

Number of actors. The participants rarely described events with multiple agents ($M = .26$), but as is shown in Figure 3, they did so more often with coarse instructions [$F(1,17) = 9.55, p < .01$]. This suggests that overall, the participants tended to focus on the actions of single agents, likely because monitoring multiple agents places more demands on attention. But perceived coarse events were more likely to involve multiple agents, suggesting that at the coarse level, the observers were better able to identify interactions among the agents. No other effects were reliable. This difference between coarse and fine descriptions is consistent with prior findings from Oatley and Yuill's (1985) study, that observers of abstract events describe coordinated interactions only once they can integrate information about several events across time. It thus makes sense that multiagent, more social interpretations emerge more often when events are thought about at a coarse level.

Table 2
Verbs in the Novel and Familiar Conditions
That Comprised at Least 1% of All Verbs

Novel Condition	Percent of Total Verbs	Familiar Condition	Percent of Total Verbs
Move	34.40	Move	17.91
Go	16.80	Go	12.43
Stop	9.52	Be	12.37
Close	7.50	Chase	9.75
Open	4.43	Come	7.69
Chase	4.10	Run	5.18
Come	3.78	Leave	4.70
Start	3.67	Find	4.28
Leave	3.23	Try	3.00
Spin	2.22	Open	2.61
Be	1.40	Hide	2.01
Follow	1.10	Talk	1.88
		Close	1.67
		Look	1.53
		Break	1.34
		Enter	1.18
		Start	1.15
		Push	1.13

Hierarchical Interpretation

The descriptions can also give insight into the hierarchical organization of the events. For routine, everyday events such as making a bed, the use of nouns and verbs—objects and actions—sharply separated coarse and fine units (Zacks et al., 2001). Coarse units evoked more specific nouns and more generic verbs than did fine units (e.g., *she put the sheet on*), whereas fine units evoked more general nouns and more specific verbs than did coarse units (e.g., *she tucked it in*; Zacks et al., 2001). The present events, however, portrayed motion paths, akin to whole-body motion, rather than actions (verbs) on objects (nouns). An appropriate index of hierarchical organization for whole-body motion should be the relative differentiation of verbs at coarse and fine levels. For example, a set

of fine events might be described as “the square moved around the corner; the square opened the door; the square moved inside.” At the coarse level, these fine events might be grouped together with a different verb, as in “the square chased the circle into the box.”

A difference index for measuring hierarchical interpretation was created by first computing the total number of *different* verb types (not tokens) produced by each participant for coarse and fine segments. The number of verb types used for both coarse and fine descriptions was divided by the total number of verb types, and this ratio was subtracted from one. For example, if a participant used 16 different verbs, 9 of which were used in both coarse and fine unit descriptions, the difference index would equal .44. Observers who use many different verbs for fine and coarse segments will have a higher difference index.

As is shown in Figure 4, observers in the familiar condition had a higher average difference index, suggesting that they differentiated actions at coarse and fine levels of description more than did observers in the novel condition [$F(1,31) = 9.51, p < .01$]. Familiarity thus increased the hierarchical nature of the observers’ intentional interpretations. Film direction had no effect on this measure [$F(1,31) = 0.03, p = .87$].

Interpretation Complexity

With familiarity, interpretations also became more complex and differentiated. For each participant, an index of interpretation complexity was computed by dividing the total number of different verbs by the total number of utterances.¹ This index was calculated separately for coarse and fine levels and then averaged for each participant. Participants who have richer, more complex interpretations of the films should also perceive more subtlety in the meaning of the events, leading them to describe the film using a larger variety of verbs. Confirming this prediction, the participants viewing familiar films produced a more diverse set of verbs ($M = .68, SEM = .03$) than did the participants viewing novel films ($M = .58, SEM = .04$)

[$F(1,31) = 5.03, p < .05$]. Film direction had no effect on interpretation complexity [$F(1,31) = 0.05, p = .83$].

Event Segmentation

The participants interpreted the events differently depending on whether the events were presented forward or backward, once or five times. Were the changes in interpretations accompanied by changes in event segmentation and organization? In the next sections, we will examine the effects of film direction, familiarity, and describing on segment boundaries, degree of hierarchical organization in the segments, and number of perceived segments. To preview the findings, despite the changes in event interpretations with familiarity and film direction, there was remarkable consistency in the organization of perceived events and the locations of event boundaries. Consistent with previous research, the number of perceived units did decrease when events were familiar or described.

Number of Perceived Units

Previous studies have shown that when events are more ambiguous or unfamiliar, segmentation becomes finer (Newton, 1973; Newton et al., 1977; Vallacher & Wegner, 1987; Wilder, 1978a, 1978b; Zacks et al., 2001). Consistent with these studies, the participants in the novel condition segmented the films into significantly more units than did the participants in the familiar condition [$F(1,74) = 13.84, p < .001$; see Figure 5]. Furthermore, the participants in the describe condition identified fewer events than did the participants in the silent condition [$F(1,74) = 8.92, p < .01$; see Figure 5]. It is possible that the decrease in the number of segments in the describe condition was due to interference from the time pressure of both segmenting and describing. If this is true, familiarity with the films should *reduce* any interference between the two tasks of describing and segmenting. Supporting this possibility, describing led to a slightly *smaller* decrease in the number of segments in the familiar condition, as is shown in Figure 5, although the interaction between describing and familiarity was not

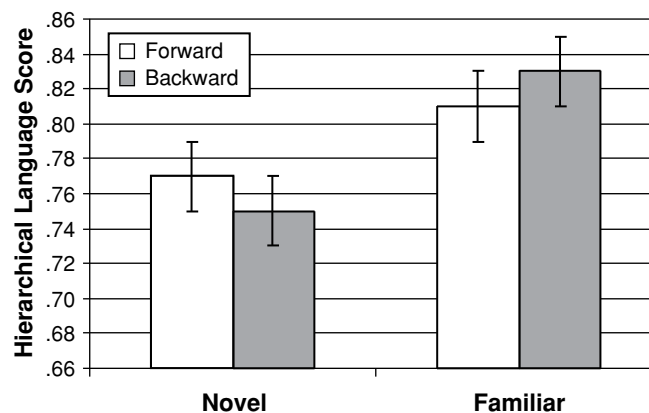


Figure 4. Mean difference index (ratio of nonoverlapping verbs to total verbs) and standard errors as a function of film direction and experience.

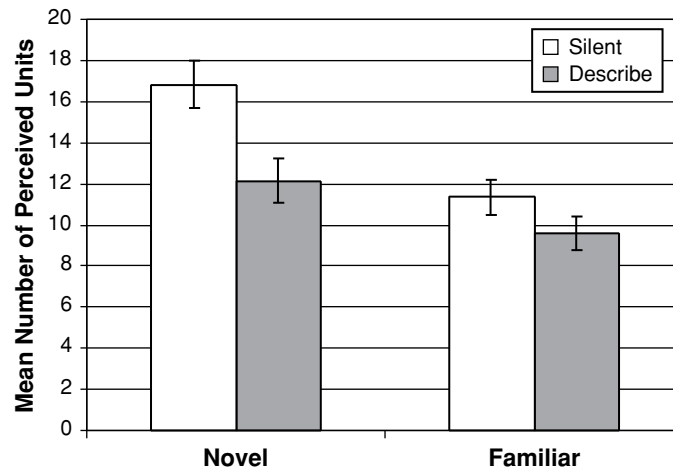


Figure 5. Mean number of perceived units and standard errors as a function of language condition (silent or describe) and experience (novel or familiar).

significant [$F(1,74) = 1.90, p = .72$]. However, time pressure cannot explain the drop in the number of segments for familiar events, in comparison with novel events, a larger effect than the effect of describing. Previous findings offer a plausible explanation for the decrease in segments for familiar films: that familiarity elicits top-down information, yielding a focus on goals during segmentation (Zacks et al., 2001). A focus on goals leads observers to group smaller events into more meaningful wholes.

The decrease in perceived units for familiar events suggests that forward films should elicit fewer units than do backward films. However, the opposite was true, although the effect was small. Forward films elicited slightly more perceived events than did backward films [13.32 to 12.49, $SEM = 0.65$; $F(1,74) = 4.26, p < .05$]. This result is surprising, given the opposite effects of familiarity. It is possible that when events are truly baffling, observers focus more globally, rather than focusing on details (Markus, Smith, & Moreland, 1985). Film direction did not interact with any other factors.

Finally, consistent with the instructions, the participants identified more units during fine segmentation ($M = 18.45, SEM = 0.90$) than during coarse segmentation ($M = 7.36, SEM = 0.40$) [$t(77) = -16.44, p < .001$].

Hierarchical Alignment

Previous findings suggested that events are segmented more hierarchically when observers identify the function or goals underlying sequences of events (Newton, 1973; Vallacher & Wegner, 1987; Zacks et al., 2001). One consequent question is whether varying event interpretations through film direction and familiarity also influenced alignment between coarse and fine event boundaries. The methods for analyzing hierarchical alignment will be detailed in the Appendix.

Hierarchical alignment was above chance for all the conditions, as is described in the Appendix. Furthermore,

backward events ($M = .35, SEM = .04$) had the same mean alignment as did forward events ($M = .34, SEM = .03$) [$F(1,73) = 0.04, p = .85$]. This finding is surprising, given that forward films were interpreted as more intentional than were backward films, and suggests that perceiving intentions is not necessary to perceive hierarchical event structure. But given that forward films were not interpreted more *hierarchically* than backward films, similar alignment for forward and backward events makes sense. Event familiarity, on the other hand, did influence hierarchical interpretations. Yet alignment scores were similar for the novel ($M = .32, SEM = .04$) and the familiar ($M = .36, SEM = .04$) conditions [$F(1,73) = 0.61, p = .44$]. Finally, contrary to previous findings by Zacks et al. (2001), the participants who described did not have higher hierarchical alignment ($M = .29, SEM = .05$) than did the participants who were silent ($M = .35, SEM = .04$) [$F(1,73) = 1.33, p = .25$]. No interactions were significant.

The participants in the present study identified event segments that were hierarchically aligned; coarse unit boundaries coincided with fine unit boundaries. The extent to which event segments were hierarchically aligned did not vary across conditions, despite differences in interpretations across conditions. How can we explain consistency in hierarchical alignment regardless of interpretation?

There are at least three possibilities. First, hierarchical alignment might not be a sensitive enough measure to detect differences in hierarchical perception. This possibility seems unlikely, given that other studies (e.g., Dowell, Martin, & Tversky, 2004; Zacks et al., 2001) have shown significant effects using this measure. Nevertheless, we also assessed hierarchical structure using a measure developed by Hard et al. (2006). This measure, called *enclosure*, is based on the asymmetry of the relationship between fine breakpoints and corresponding coarse ones: Coarse breakpoints often occur slightly after their closest fine breakpoint; when they occur before, they tend to

correspond to an event different from that captured by the corresponding fine unit. Accordingly, the enclosure measure pairs up coarse and fine breakpoints and assesses the extent to which coarse breakpoints occur after, or *enclose*, their closest fine breakpoint. This measure is a good predictor of hierarchical descriptions for events and also tends to be higher for participants who segment while describing and who are explicitly instructed to look for hierarchical relations among actions (Hard et al., 2006; Lozan, Hard, & Tversky, in press). Yet for these abstract events, this measure yielded no differences in hierarchical segmentation based on film direction, experience, or describing.²

Second, task demands might have encouraged the participants to identify hierarchically organized segments, regardless of their interpretations. This, too, seems unlikely; it requires that observers remember on the second viewing where they segmented during first viewing and that they segment at the same places as before, either fewer or more times, depending on order. There is no evidence that people adopt that strategy even for routine everyday events, which are easy to interpret and easier to remember than the events presented here (Zacks et al., 2001).

Third and most likely, hierarchical segmentation might have a physical basis, so that observers can identify hierarchically organized events even when they cannot interpret them. If this is the case, observers should identify approximately the same units despite differences in interpretation. Under this argument, the physical bases of segmentation should be knowable, and further analyses will demonstrate that they are.

Agreement on Event Boundaries

Do participants across conditions agree on event boundaries? To determine whether there was agreement across conditions, each film was divided into eighty-five 1-sec bins, and the total number of participants who indicated a coarse or fine breakpoint in each bin was determined. Pearson correlations were computed for segmentation between forward and reforwarded backward films. They were also computed between films viewed once and films viewed five times.

Forward and backward events. If event segmentation is determined by salient changes in physical movement and these changes are perceptually similar both forward and backward, observers of forward films should select the same breakpoint locations as observers viewing the backward version of the same film. In fact, the correlations for both films and familiarity conditions were high, as is shown in Table 3. The weakest correlation was for hide-and-peek in the familiar condition ($r = .20, p = .06$). These correlations are all the more impressive given the noisiness of these data. The observers often reported having “missed” marking event boundaries or being late to respond to them, factors that would reasonably reduce observer agreement. These results offer evidence that the observers identified similar event boundaries with similar frequency, regardless of event direction.

Table 3
Correlation Coefficients Comparing the Number of Breakpoints in Each 1-sec Interval for Forward Versus Backward Films and Novel Versus Familiar Films

Film	Forward Versus Backward		Novel Versus Familiar	
	Novel	Familiar	Forward	Backward
Chase	.31**	.49*	.59*	.69*
Hide-and-peek	.36**	.20***	.60*	.48*

Note—Correlations were calculated independently for each film and condition. * $p < .001$. ** $p < .01$. *** $p < .10$.

Novel and familiar events. Do observers choose similar event boundaries despite differences in familiarity? As Table 3 shows, high, reliable correlations were found between the novel and the familiar conditions. This shows that similar breakpoints were selected irrespective of experience, despite striking differences in how the events were interpreted. Figure 6 shows the similarity between breakpoint distributions in the two familiarity conditions graphically, with data from the novel condition plotted as negative, to show the symmetry in the breakpoint distributions.

The Physical Basis of Event Structure

So far, the data indicate that goal-based event schemas play a minimal role in detecting event structure. The participants selected similar, hierarchically organized events regardless of whether they interpreted those events as intentional actions or physical movements. These findings suggest that event structure has a physical basis that observers can exploit *before* building event interpretations. Consistent with this explanation, the participants who segmented the events on the first viewing selected the same event boundaries as did the observers who had seen the events multiple times and constructed a narrative for them.

What is the physical basis of event structure? Newton et al. (1977) argued that event boundaries for human action coincide with bursts of change in the body position of the actor. Perhaps observers of abstract events also use bursts of change in the positions of the geometric “agents” to detect event boundaries. Coarse and fine event boundaries might differ in the magnitude of these bursts. Alternatively, coarse and fine event boundaries might coincide with qualitatively different changes. Evidence from Zacks et al. (2001), for example, suggests that for some everyday activities, observers individuate coarse-grained events by changes in the objects being acted on and fine-grained events by changes in the action being performed on a given object.

To test the possibility that observers use physical cues to detect hierarchical event structure, the *chase* and *hide-and-peek* films were coded for several types of movement changes and subjected to two analyses. The first analysis examined quantities of movement changes, comparing breakpoints with nonbreakpoints and coarse breakpoints

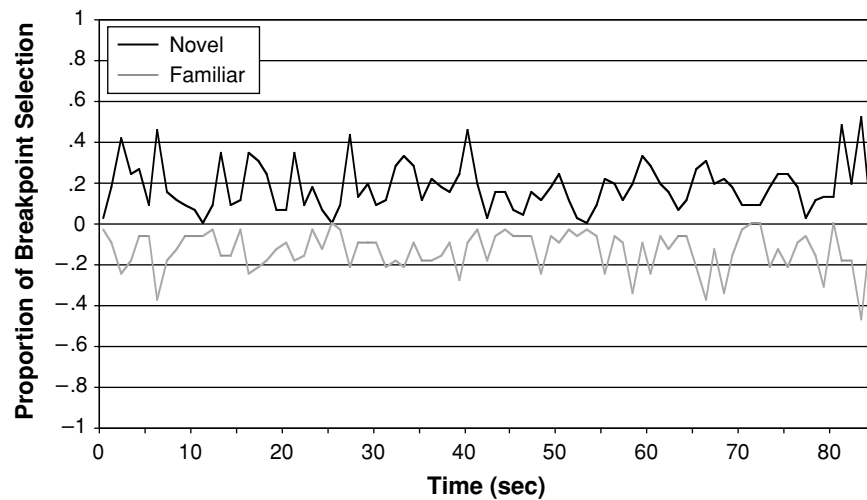


Figure 6. Proportion of times that a 1-sec bin for chase (forward) was selected as a fine or coarse breakpoint. The proportions for the participants in the familiar condition are shown as negative, to facilitate visual comparison with those for the participants in the novel condition.

with fine ones. The second analysis compared the types of movement changes that cue coarse and fine breakpoints, using stepwise linear regression. To preview the findings, breakpoints involved more movement changes than did nonbreakpoints, and coarse breakpoints involved more movement changes than did fine breakpoints. Coarse and fine breakpoints did not differ significantly in the types of changes that cued them.

Coding Procedure

A trained coder viewed each film repeatedly and, for each 1-sec interval, coded the values of several binary movement variables described below (1 if a movement change was present, 0 if it was absent). These movement variables were coded separately for each of the three geometric “agents” in each film.

Start. Within a given interval, an agent started to move if its geometric center was initially stationary and then changed location. Rotations were not considered to be starts when the geometric center of the agent remained stationary.

Stop. An agent stopped moving if its geometric center was moving and then became stationary.

Change direction. An agent changed its direction if the direction in which its geometric center was moving changed abruptly. Slow, gradual changes in direction did not count as direction changes.

Turn. An agent turned if it changed its orientation by less than 360° .

Rotate. An agent rotated if it completed at least one full 360° rotation.

Contact object. An agent contacted another object if it appeared to physically touch another object.

Change speed. An agent changed speed if there was an obvious and abrupt change in its speed. A slow, gradual

increase or decrease in speed did not count as a change in speed.

Quantities of Movement Change

A breakpoint histogram was created for each participant by binning that participant’s keypress responses into 1-sec intervals. Because each of these 1-sec intervals had been coded for different types of movement changes, we could then determine, for each participant, the average number of total movement changes (summed across the three geometric agents) that occurred during nonbreakpoint intervals, fine breakpoint intervals, and coarse breakpoint intervals.³ These scores were submitted to a repeated measures ANOVA with interval type (nonbreakpoint, fine breakpoint, or coarse breakpoint) and event direction as factors (collapsing across experience and language conditions, which had no effects). There was a main effect of interval type on the number of movement changes [$F(2,154) = 139.04, p < .001$]. Film direction had no effect and did not interact with the effect of interval type. So, two planned contrasts compared breakpoints (collapsed across coarse and fine) with nonbreakpoints and coarse breakpoints with fine breakpoints.

Breakpoints corresponded to significantly more movement changes ($M = 3.16, SEM = .04$) than did nonbreakpoints ($M = 2.32, SEM = 0.012$) [$F(1,77) = 413.21, p < .001$]. Coarse and fine breakpoints did not differ reliably [$F(1,77) = 0.38, p = .52$]. However, because fine and coarse breakpoints often coincide by virtue of hierarchical segmentation, the fine breakpoints in this analysis also contained coarse breakpoints, potentially diluting any difference between coarse and fine breakpoints. To counter this problem, a second analysis compared fine breakpoints that fell closest to coarse event breakpoints (external fine breakpoints) with pure fine breakpoints, those

that fell within coarse events (internal fine breakpoints). In fact, external fine breakpoints were characterized by more movement changes ($M = 3.38, SEM = 0.06$) than were internal fine breakpoints ($M = 2.95, SEM = 0.053$) [$F(1,77) = 26.01, p < .001$]. Thus, magnitude of change does differentiate coarse and fine breakpoints.

Further analysis of internal and external fine breakpoints showed that breakpoint intervals corresponded to dramatic bursts in the number of movement changes, relative to preceding intervals. The quantity of movement changes for breakpoint intervals (internal and external fine) was compared with the quantity of movement changes for the prebreakpoint intervals immediately preceding them⁴ and with nonbreakpoints. The internal and external breakpoints were characterized by more movement changes than were the prebreakpoint intervals [$F(1,77) = 146.41, p < .001$]. As is shown in Figure 7, this effect interacted with breakpoint type and was stronger for external fine breakpoints, those that coincided with coarse unit boundaries, than for internal fine breakpoints [$F(1,77) = 5.03, p < .05$]. As Figure 7 also shows, there was a difference between prebreakpoint intervals and other nonbreakpoints. Specifically, at prebreakpoint intervals, there was a slight but significant increase in the number of movement changes, relative to nonbreakpoints in general [$F(1,77) = 18.52, p < .001$]. This suggests that movement changes begin to increase *before* breakpoints and then jump dramatically at breakpoints. This might provide observers with a means of anticipating impending event boundaries.

Qualities of Movement Change

In natural human action, different types of movement changes might distinguish coarse event boundaries from fine boundaries. Consider the making of a bed. At the coarse level, the actor acts on different objects (i.e., sheets, comforter, pillows). Each of these events can be individuated by a predictable sequence of movements in the actor: releasing one object, orienting the head and body toward the next object, and then moving and contacting that ob-

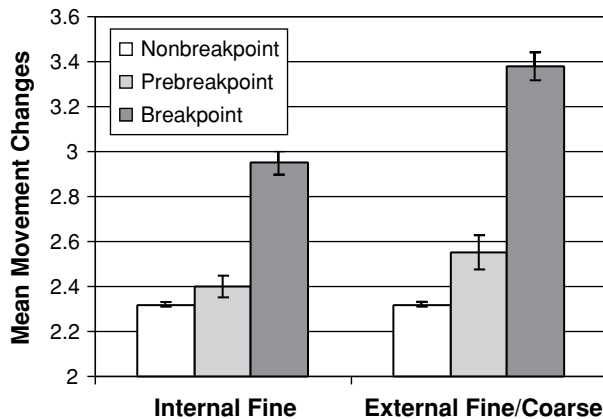


Figure 7. Mean number of movement changes and standard errors for nonbreakpoints, prebreakpoints, and breakpoints. Internal and external fine breakpoints are depicted separately.

Table 4

Grain	R	SE	Most Predictive Changes
Coarse	.19	3.59	start, turn
Total fine	.34	5.03	start, contact object, stop, change direction
External fine	.24	3.52	stop, start, change speed
Internal fine	.33	3.01	start, contact object, change speed

Note—Changes are listed in the order in which they entered the stepwise regression, up to the last feature that accounted for at least 1% of incremental variance.

ject (Baldwin & Baird, 1999). At the fine level, the actor performs different actions on the same object. Each of these events might be individuated by the actor’s hand releasing and then recontacting the same object. How might observers use different types of movement cues to segment abstract events on different grains?

To test the possibility that coarse and fine event boundaries coincide with different movement cues, all of the participants’ coarse and fine breakpoint histograms for each film and film direction were compiled, yielding two group-level breakpoint histograms (coarse and fine). Each histogram represented the number of participants indicating a breakpoint within a given 1-sec interval. The coded movement changes⁵ were then entered as predictors into a forward stepwise regression to explain the variance in coarse and fine breakpoints across intervals. It should be noted that the bins were long enough to allow more than one kind of movement change. A similar analysis compared external and internal fine breakpoints.

The results are displayed in Table 4. There was no evidence that coarse and fine breakpoints are cued by qualitatively different changes. Similar cues predicted coarse breakpoints, external fine breakpoints, and internal fine breakpoints. There were some slight differences in the order that the predictors entered the different regressions, but these differences were not large enough to be meaningful. What was meaningful was that *starts* and *stops* were the most predictive cues for all levels of breakpoints. This suggests that observers segment not only on the basis of bursts in movement change, but also on the basis of pauses in the actions of the geometric agents, just as they do in natural human action (Barker & Wright, 1955). Furthermore, *starts* for these 2-D events are similar to the initiations of movement that live actors make toward an object that they plan to act on. In these films, the “objects” that the agents interact with are typically other agents. Thus, breakpoints corresponded not only to bursts in the number of features changing, but also to goal-relevant cues, such as initiation of movement.

DISCUSSION

People perceive structure in everyday events, segmenting them into parts and organizing the parts into larger parts. The structure perceived in observed events reflects the structure of event schemas, or explicit knowledge of goals and subgoals. Do people require event schemas to detect structure in observed events? Previous studies have

suggested that schematic knowledge about the goals or functions underlying sequences of events might be necessary to group smaller events into larger ones.

This question was addressed directly by asking participants to segment abstract events that consisted of geometric figures moving around an environment according to a chase or a hide-and-seek scenario. The participants' ability to interpret the events by using top-down knowledge was manipulated in two ways. First, the events were shown forward or backward in order to examine the effects of prior event knowledge, such as scripts for chase or hide-and-seek, on the participants' ability to segment the events. Second, the participants' familiarity with these specific stimuli was manipulated. Half of the participants described as they segmented, providing a check on people's interpretations of the events. As was predicted, the participants interpreted forward events as more intentional than backward events and familiar events as more intentional than novel events. Furthermore, the participants viewing familiar events differentiated their descriptions of fine and coarse segments, suggesting that they had arrived at deeper hierarchical interpretations.

Irrespective of interpretations, the participants selected hierarchically organized segments with essentially the same boundaries, or breakpoints. These findings suggest that for these events, physical changes in the actions subserve event segmentation and hierarchical organization. They imply that goal-based event schemas are not required to detect event structure. Instead, physical information constrains segmentation, leading observers to identify the same breakpoints even when their ongoing event representations differ.

A point worth considering is that the abstract events used in the present study were generated from intentional scripts. It is possible, even likely, that these events, even when shown backward, are interpreted more intentionally than randomly generated movements would be. Furthermore, the practice film might have encouraged the participants to look for intentionality in the test films, because it was based on the intentional script of freeze tag. The scripted intentionality in the events, and the participants' expectations of intentionality, may have influenced hierarchical segmentation. Some recent evidence suggests that hierarchical segmentation was a consequence neither of inherent intentionality in the films nor of expectations of intentionality. In a study by Zacks (2004), participants segmented videos of geometric shapes, moving randomly or as a consequence of people's playing a video game. Whether the movements were *truly* intentional or random did not affect the degree of hierarchical segmentation. Furthermore, whether the participants were *told* that the movements were truly intentional or random did not affect the degree of hierarchical segmentation (J. M. Zacks, personal communication, June 8, 2005). Together, these results support the interpretation that goal-based schemas are not required to perceive hierarchical structure in event sequences.

Further evidence that hierarchical encoding was not merely a consequence of inherent intentionality in the

films or expectations of intentionality comes from the fact that components of motion were tightly tied to coarse and fine levels of segmentation. Breakpoints corresponded to bursts of movement changes, indicating that the observers perceived event structure by monitoring discontinuities in the flow of information. Coarse and fine breakpoints differed in the magnitude of these bursts: Coarse breakpoints involved more changes than did fine breakpoints. Even in randomly generated events, changes in quantities of movement are inevitable. Observers may use the relative magnitude of these changes to organize even random events hierarchically. The simple motion cues that predicted breakpoints suggest that observers are especially attentive to movement changes that are goal relevant. Specifically, the observers identified breakpoints when movement was initiated, a cue that corresponds to new goals in natural human action (Baldwin & Baird, 1999; Barker & Wright, 1955).

Do the findings presented here generalize to the perception of natural human behavior? Breakpoints in human activity do correspond to bursts of change in body position (Newtson et al., 1977), suggesting that a physical basis for segmentation can be detected in the absence of event schemas, even for actions entailing articulated movements of the body, and not just paths of motion. Previous studies did not compare the magnitudes of these bursts at coarse and fine event boundaries. It is possible that structure in human activity is similarly cued by bursts of different magnitude. It is also possible, however, that the differences found here between coarse and fine breakpoints are specific to multiagent events, such as the ones depicted in these films. Perhaps more movement features changed at coarse breakpoints because these breakpoints more often involved multiagent events, as is suggested by the verbal descriptions.

An essential difference between the events in the present study and those in natural human action is that the present films portrayed paths of motion, rather than articulated actions on objects. Intentions and motion might be more tightly coupled in motion paths than in hand and arm movements, making it easier for observers to extract structure by monitoring movement changes. Hand manipulations in such events as making a bed are far more subtle, complex, and variable than the events in the present study, with some actions more closely linked to goals and others incidental. In some cases, incidental or less central changes in hand and body motions might be more salient than those closely linked to goals. Conceptual knowledge may be helpful in distinguishing which bursts in movement change are relevant to goals.

In the present study, conceptual knowledge did not change overall patterns of segmentation but did lead to coarser segmentation, consistent with previous findings (Newtson, 1973; Newtson et al., 1977; Wilder, 1978a, 1978b; Zacks et al., 2001). Whether familiar or novel, coarse events were interpreted more conceptually than were fine events. Coarse-level descriptions more often referred to intentions and to multiagent events; fine-level descriptions more often referred to motion paths. Com-

bined, these results suggest that goal-based interpretations can influence event segmentation, even if they are not required. When observers have intentional interpretations, they choose more selectively from the set of possible breakpoint locations, leading to coarser segmentation. Perceived intentions aggregate events into larger wholes by conferring relations on them.

The parsing of events, then, seems to lead to intentional knowledge, as well as resulting from it. The present findings show that discontinuities in movement information enable hierarchical event segmentation in the absence of event schemas. Because these discontinuities correspond to meaningful changes in intentions, they may facilitate construction of hierarchical, goal-based event schemas. This point is critical to understanding the development of intentional understanding in infants and is consistent with findings that infants who lack adults' intentional knowledge can, nevertheless, parse behavior as adults do (Baldwin et al., 2001). Changes in movement features might provide infants with a simple means of segmenting behavior into meaningful units. Once these units are identified, they can be used to build explicit knowledge about goals and intentions. This knowledge, in turn, may affect subsequent segmentation and hierarchical organization.

The present research suggests that the same process occurs for adults experiencing abstract events that are difficult to interpret, especially ones that are novel or shown backward to disrupt their causal sequencing. Observers can use physical information to identify organized units within an event sequence and then use these units to build interpretations. Initially, these interpretations are movement based and shallow. With experience, interpretations deepen and become tied to knowledge about agents' goals and subgoals. This outcome is reminiscent of Gibson and Gibson's (1955) observations on perception. Like wine tasters, observers of events come to perceive more in what is out there (perceiving intentionality and hierarchical structure, however, was not what Gibson & Gibson meant). Once hierarchical event schemas are formed, they can then facilitate the perception of event structure, helping observers segment and organize their experiences.

REFERENCES

- ABBOTT, V., BLACK, J. B., & SMITH, E. E. (1985). The representation of scripts in memory. *Journal of Memory & Language*, **24**, 179-199.
- BALDWIN, D. A., & BAIRD, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 215-240). Mahwah, NJ: Erlbaum.
- BALDWIN, D. A., BAIRD, J. A., SAYLOR, M. M., & CLARK, M. A. (2001). Infants parse dynamic action. *Child Development*, **72**, 708-717.
- BARKER, R. G. (1963). The stream of behavior as an empirical problem. In R. G. Barker (Ed.), *The stream of behavior: Explorations of its structure and content* (pp. 1-22). New York: Appleton-Century-Crofts.
- BARKER, R. G., & WRIGHT, H. F. (1955). *Midwest and its children: The psychological ecology of an American town*. Evanston, IL: Row, Peterson.
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177-220.
- BYRNE, R. W. (1999). Imitation without intentionality: Using string parsing to copy the organization of behaviour. *Animal Cognition*, **2**, 63-72.
- BYRNE, R. W. (2002). Seeing actions as hierarchically organized structures: Great ape manual skills. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 122-140). Cambridge: Cambridge University Press.
- CASATI, R., & VARZI, A. C. (Eds.) (1996). *Events*. Brookfield, VT: Dartmouth.
- CLEARFIELD, M. (2004). Infants' enumeration of dynamic displays. *Cognitive Development*, **19**, 309-324.
- 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, & Computers*, **25**, 257-271.
- DICKMAN, H. R. (1963). The perception of behavioral units. In R. G. Barker (Ed.), *The stream of behavior: Explorations of its structure and content* (pp. 23-41). New York: Appleton-Century-Crofts.
- DOWELL, R. E., MARTIN, B. A., & TVERSKY, B. (2004). Segmenting everyday actions: An object bias? In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (p. 1553). Mahwah, NJ: Erlbaum.
- GIBSON, J. J., & GIBSON, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, **62**, 32-41.
- HARD, B. M., LOZANO, S. C., & TVERSKY, B. (2006). Hierarchical encoding of behavior: Translating perception into action. *Journal of Experimental Psychology: General*, **135**, 588-608.
- 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.
- LASHLEY, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York: Wiley.
- LOZANO, S. C., HARD, B. M., & TVERSKY, B. (in press). Perspective-taking promotes action understanding and learning. *Journal of Experimental Psychology: Human Perception & Performance*.
- MARKUS, H. [R.], SMITH, J., & MORELAND, R. L. (1985). Role of the self-concept in perception of others. *Journal of Personality & Social Psychology*, **49**, 1494-1512.
- NEISSER, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.
- NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- NEWTON, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality & Social Psychology*, **28**, 28-38.
- NEWTON, D., & ENGQUIST, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, **12**, 436-450.
- NEWTON, D., ENGQUIST, G., & BOIS, J. (1977). The objective basis of behavior units. *Journal of Personality & Social Psychology*, **35**, 847-862.
- OATLEY, K., & YUILL, N. (1985). Perception of personal and interpersonal action in a cartoon film. *British Journal of Social Psychology*, **24**, 115-124.
- SHARON, T., & WYNN, K. (1998). Individuation of actions from continuous motion. *Psychological Science*, **9**, 357-362.
- TVERSKY, B., & HEMENWAY, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, **113**, 169-193.
- VALLACHER, R. R., & WEGNER, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, **94**, 3-15.
- WAGNER, L., & CAREY, S. (2003). Individuation of objects and events: A developmental study. *Cognition*, **90**, 163-191.
- WILDER, D. A. (1978a). Effect of predictability on units of perception and attribution. *Personality & Social Psychology Bulletin*, **4**, 281-284.
- WILDER, D. A. (1978b). Predictability of behaviors, goals, and unit of perception. *Personality & Social Psychology Bulletin*, **4**, 604-607.
- WOODWARD, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, **69**, 1-34.

- WYNN, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, 7, 164-169.
- ZACKS, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28, 979-1008.
- ZACKS, J. M., & TVERSKY, B. (2001). Event structure in perception and cognition. *Psychological Bulletin*, 127, 3-21.
- ZACKS, J. M., TVERSKY, B., & IYER, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130, 29-58.

NOTES

1. The total number of verbs was divided by the total number of utterances to account for the possibility that participants produce more verbs simply because they are producing more descriptions overall.
2. There were no significant differences in enclosure scores for forward ($M = .59$, $SEM = .03$) versus backward ($M = .64$, $SEM = .03$) events [$F(1,74) = 2.15$, $p = .15$], for familiar ($M = .60$, $SEM = .03$)

versus novel ($M = .63$, $SEM = .02$) events [$F(1,74) = 0.61$, $p = .44$], or for participants who described versus ones who were silent [$M = .61$, $SEM = .03$, for both; $F(1,74) < 0.001$, $p > .99$].

3. We expected that the participants might take a moment to notice that a change burst had occurred. In calculating the average total movement features for each interval type (nonbreakpoint, fine breakpoint, and coarse breakpoint), we thus summed the movement changes within the 1-sec interval that the participant pressed the space bar and in the 1-sec interval preceding it. For example, if a participant indicated a fine breakpoint in Interval 36, we counted the total number of movement changes in Interval 36 plus those in Interval 35.

4. Prebreakpoint intervals corresponded to a 2-sec interval before a breakpoint interval. For example, if a participant indicated a breakpoint in Interval 36, the prebreakpoint interval would include the total movement changes in Interval 33 plus those in Interval 34.

5. For simplicity, we created new binary variables that indicated whether *any* of the three geometric agents had made a given movement change.

APPENDIX

Before comparing hierarchical alignment across conditions, we first determined whether fine and coarse unit boundaries aligned more often than chance. We used the discrete and continuous analyses developed by Zacks et al. (2001) and described below.

Discrete Method

Because each film was slightly over 84 sec long, each film was divided into eighty-five 1-sec bins. For each participant, a bin was considered a *breakpoint* if it contained one or more space bar presses, and a bin was considered to be an *overlap* if it qualified as a breakpoint for both the coarse and the fine unit conditions. If the boundaries of fine unit breakpoints are unrelated to the boundaries of coarse unit breakpoints, the expected number of overlaps for each participant would equal the product of the probability that a bin contains a coarse breakpoint and the probability that a bin contains a fine breakpoint, multiplied by the total number of bins. This formula provided a null model with which to compare the observed number of overlaps for each participant.

Despite the abstractness of the stimuli, there were far more overlaps between coarse and fine boundaries than would be predicted by chance, both for forward films [observed overlaps, $M = 3.18$, $SEM = 0.34$; expected overlaps, $M = 1.87$, $SEM = 0.17$; $t(77) = 6.11$, $p < .001$] and for backward films [observed overlaps, $M = 2.88$, $SEM = 0.29$; expected overlaps, $M = 1.67$, $SEM = 0.18$; $t(77) = 7.47$, $p < .001$].

Continuous Method

Because a discrete analysis depends on the arbitrary choice of discrete bin size, we also adopted the continuous analytic procedure developed by Zacks et al. (2001). In this analysis, each space bar press is considered a breakpoint. For each coarse breakpoint, the distance to the nearest fine breakpoint was calculated. We averaged these distances to determine the mean distance for each participant. We determined a null model for the expected distance between coarse and fine breakpoints by letting $F = \{f_1, f_2, \dots, f_{\text{Fine}}\}$ be the set of all fine breakpoints for a given participant in milliseconds. We then used the following formula:

$$\text{AvgDist}_0 = \frac{\frac{f_1^2}{2} + \sum_{i=1}^{i=\text{Fine}-1} \left[\frac{f_{i+1} - f_i}{2} \right]^2}{f_{\text{Fine}}}$$

The basis for this formula is described in more detail in Zacks et al.

A log transform was computed for the observed and expected average distances for each participant in order to reduce positive skewness. As is shown in Figure A1, we found an average distance between each coarse break-

APPENDIX (Continued)

point and its closest fine breakpoint that was smaller than that predicted by the null model, both for forward films [$t(77) = -6.85, p < .001$] and for backward films [$t(77) = -7.37, p < .001$].

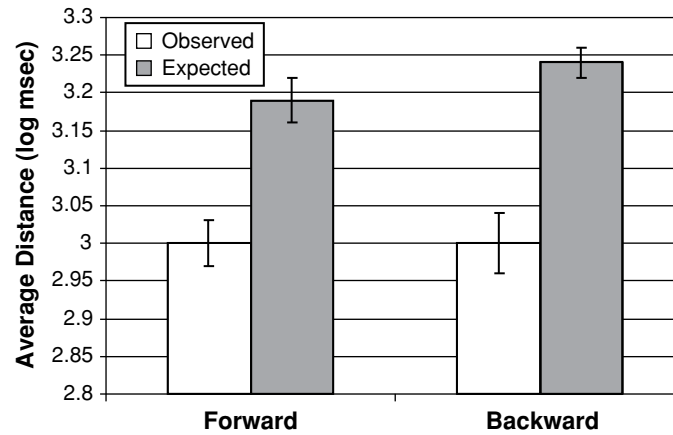


Figure A1. Observed and expected average distances (in log milliseconds) and standard errors between coarse unit boundaries and their nearest fine unit boundaries for forward and backward films.

To determine whether alignment varied across conditions, the degree of hierarchical alignment was computed on the basis of the continuous analysis described above.

Zacks et al. (2001) used the difference between expected and observed average distance as their alignment score. But given that participants vary in the number of perceived units, which influences expected average distance, two participants can have equal observed average distances but look falsely different in alignment score. Using the ratio of the difference (expected – observed average distance) to the expected average distance corrects for this problem. We calculated these ratios for each participant and film. A mixed factorial ANOVA was conducted on these alignment scores, with film direction and language condition as factors.

We found the same results when we used Zacks et al.'s (2001) original alignment scores and when we analyzed alignment scores derived from the discrete analysis and from the logarithmic transforms of the continuous data.