Segmenting Ambiguous Events

Bridgette A. Martin (martin@psych.stanford.edu)
Department of Psychology, Bldg. 420 Jordan Hall
Stanford, CA 94305 USA

Barbara Tversky (bt@psych.stanford.edu)
Department of Psychology, Bldg. 420 Jordan Hall
Stanford, CA 94305 USA

Abstract

Everyday events, such as making a bed, are segmented hierarchically, with the coarse level punctuated by objects or object parts and the fine level by articulated actions on objects. Here we examine segmentation of events involving abstract, ambiguous motion paths of several geometric figures, viewed once or five times. Segmentation was hierarchical for both; however after one viewing, events were interpreted as movements whereas after five viewings, they were interpreted as intentional actions. Fewer (but the same) segments were identified after five viewings. Experience did not affect segment boundaries but did affect segment interpretation, shifting from bottom-up to top-down.

Introduction

The physical world is in constant flux. From fields of grass to bustling city streets, the world is characterized by matter in motion, changing speed and direction, interacting with other matter, rarely at rest. Comprehending the world entails detecting and understanding such change. To accomplish this, the mind must successfully parse continuous flux into meaningful *event*s—changes in space and time that are perceived as bounded units, with beginnings, middles, and ends (e. g., Newtson & Engquist, 1976; Zacks, Tversky, & Iyer, 2001).

Though human understanding of events encompasses both the animate (e.g., a person setting a table) and inanimate (e.g., a hurricane), research has focused on segmentation of human behavior. Parsing the behavior stream confers coherence on human movement and allows decomposition into learnable sequences, facilitating comprehension, planning, and acquisition of new skills (Zacks & Tversky, 2001). Parsing behavior is further believed to shape causal reasoning and intentional understanding (Baldwin, Baird, Saylor, & Clark, 2001).

To uncover the bases of event perception, Barker (1963) and Newtson (1973) asked observers to segment actual behavior or films of behavior into natural units. Observers agree with themselves and with others on unit boundaries, called *breakpoints* (Dickman, 1963; Newtson & Engquist, 1976, Zacks, et al., 2001). Recent work has shown that for highly schematic events, like making a bed or assembling an object, event perception is hierarchical; that is, when observers segment into the largest and smallest units that are meaningful, the boundaries of the larger units coincide with those of the smaller units (Zacks, et al., 2001). Observers' descriptions of event segments reveal that these everyday events are parsed according to actions on objects. At the coarse level, events are segmented by actions on different objects or large object parts; at the fine level, events are segmented by different actions on the same object (Zacks, et al., 2001). This finding meshes nicely with evidence that for infants' event perception, contact with objects plays a critical role (Baldwin & Baird, 1999; Woodward, 1998).

More generally, breakpoints are characterized by large changes in physical action components as well as attainments of goals or subgoals. This implies that both bottom-up and top-down information contribute to event segmentation and are correlated—as in object categories—so that bottom-up physical information can serve as a bridge to conceptual information (Rosch, 1976; Tversky & Hemenway, 1984). Top-down information about goals and intentions depends more on event familiarity than does bottom-up information about physical changes. Interestingly, when events are less familiar, more ambiguous, or less predictable, segmentation occurs at a finer level (Newtson, 1973; Vallacher & Wegner, 1987), suggesting that top-down knowledge about goals and intentions is needed to unitize smaller units into larger ones. It also indicates that bottom-up physical information is the primary factor in event segmentation when knowledge of goals is absent.

Events studied by Newtson and Zacks and their collaborators were for the most part concrete and familiar, portraying single humans enacting everyday events in rich and appropriate contexts. There have been claims that people can similarly interpret abstractly portrayed events, such as those used in the classic Heider and Simmel (1944) experiment, in terms of concrete and familiar actions. In that study, observers saw an animated film of 1 large and 2 small geometric figures moving according to an experimenter-designed script. It has been reported that people spontaneously interpret the movements of the geometric figures as human-like, intentional actions. Specifically, observers often describe the large triangle bullying the two smaller shapes: chasing them and trying to capture them.

In this animation, there are no actions on different objects or articulated actions on individual objects to guide segmentation. For these events, the actors' changing paths of motion constitute most of the activity. There are other actors, however, so actions on other actors provide one possible basis for segmentation. Changes in direction, speed, and manner of motion provide another possible basis. Because the activity in these films involves changing paths of motion, the events are more linear than events like making a bed, so an event hierarchy seems less plausible.

Our aim is to extend the methods and analysis used by Zacks and his collaborators to events that involve several actors, that are abstract and ambiguous, and that convey trajectories in space rather than manipulations of objects. What roles do top-down knowledge of goals and bottom-up information about physical changes play in segmenting these abstract events? Can these events be perceived hierarchically?

Exp 1: Abstract Events Viewed Once

The purpose of the first experiment was to investigate segmentation and interpretation of events that are ambiguous, involve more than one actor, and primarily portray paths and manners of motion. Will participants interpret these as sequences of goal-directed actions, as for previous studies of familiar and explicit events, such as washing the dishes? Will events that primarily involve changing trajectories rather than manipulations on objects be perceived to have a hierarchical structure?

Participants viewed two animations portraying motion paths of geometric figures: one based on Heider and Simmel's (1944) *chase* and the other based on *hide-and-seek.* They segmented the events at fine and coarse levels, following procedures of Newtson and Engquist (1976) and Zacks and his collaborators (Zacks, et al., 2001). Half the participants described what happened in each segment while segmenting; half only segmented. Additionally, each film was shown forward to one group and backward to another. The reasoning behind this manipulation was that if events are bounded by attainments of goals, which are unidirectional in time, viewing events backward should make goal achievement difficult to identify. This would seem to hold for events such as fertilizing a plant or making a bed. If segmentation of these abstract films relies on interpretations of goal-attainment, reversing the films should be disruptive. However, if participants segment on the basis of large physical changes in the path or manner of motion, then reversing the video may not be disruptive.

Method

Participants Fifty-one Stanford undergraduates completed the experiment in exchange for course credit.

Videos The stimuli were two, 84-s animated films created with the animation program CuriousLabs® Poser 4. Each video was two-dimensional, and portrayed three shapes, or "characters," interacting with one another and with geometric "landmarks" in the environment. One video, *chase*, was based on the script of the Heider and Simmel (1944) film, in which a large triangle bullies and chases two smaller shapes. The second video was based on *hide-and-seek*. There were two versions of each video: an original version and a perfectly reversed version.

Design The design was a 2x2x2 Mixed Factorial with film direction (forward, backward) and segmentation level (coarse, fine) varied within-subjects, and language (describe, silent) varied between-subjects. Each participant saw *chase* and *hide-and-seek*—one forward and one backward—and segmented each film twice, once into coarse and once into fine units. Both film direction and order of segmentation were counterbalanced across conditions.

Procedure Participants were told 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. Participants in the *describe* condition were additionally told that each time they pressed the space bar, they should briefly describe what happened in the segment they just observed. Half of the participants in each group were instructed to segment events into the smallest units that seemed natural and meaningful to them (fine units). The other participants were instructed to mark off events into the largest units that seemed natural and meaningful (coarse units).

Participants then viewed and segmented an 84-s practice video depicting a game of *freeze tag*. After completing the practice video, participants had the opportunity to ask questions, and then proceeded with the rest of the experiment. Videos were presented on a 21-inch, flat screen computer monitor.

After segmenting the videos, participants completed an unrelated task for 10 min. They then segmented the videos a second time in the same order, using the opposite unit-size instructions. If they had segmented by fine units the first time, then they segmented into coarse units the second time and vice versa. 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.

Results

Five participants were excluded from analysis because their number of coarse breakpoints was equal to or greater than the number of fine breakpoints for at least one video, or because for at least one video, they indicated only 1 coarse breakpoint. The data of the remaining 46 participants were analyzed for segmentation and description.

Segmentation The mean number of fine units (M = 20.71, SEM = 0.96) was significantly greater than the mean number of coarse units (M = 8.48, SEM = 0.44), $t(91) = -15.76$, $p < 0.001$), as can be seen from Figure 1.

Hierarchical structure was evaluated using the continuous analytic procedure developed by Zacks et al. (2001). In this analysis, each spacebar tap is considered a breakpoint. For each coarse breakpoint, the distance to the nearest fine breakpoint was calculated. These distances were averaged to determine the mean distance (AveDist) for each participant. To determine a null model for the expected distance between coarse and fine breakpoints, we let F = {f1, f_2, . . .,f_{fine}}, where F is the set of all fine breakpoints for a given participant in ms. We then used the 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. (2001). For each participant, a variable was created that was equal to the difference between AveDist_O and AveDist, to serve as a measure of the *degree* of hierarchical structure (larger difference equals greater alignment).

Despite the abstractness of the stimuli used in this study, a strong hierarchical alignment effect was found. As shown in Figure 2, there was a smaller average distance between each coarse breakpoint and its closest fine breakpoint (AveDist *M* = 1104.94 ms, *SEM* = 71.20 ms) than predicted by the null model (AveDist_O *M* = 1708.14 ms, *SEM* = 90.20 ms), $t(91) = -8.83$, $p < 0.001$. Intriguingly, this effect was not influenced by the direction of the film, $F(1, 30) = 0.57$, $p = 0.46$. Furthermore, contrary to previous findings by Zacks et al. (2001), describing while segmenting did not increase hierarchical alignment, $F(1, 30) = 0.04$, $p = 0.84$.

Verbal Descriptions The verbal descriptions were analyzed for intentional action, physical movement, and number of actors. Because participants often described more than one event in a single sentence, verbal descriptions were transcribed and divided into clauses. Thus each segment could receive more than one verbal description. Two coders rated a total of 1077 clauses from 18 participants. By Cronbach's Alpha, inter-rater reliability was above 0.90 for all categories.

Intentionality Clauses were coded as being "intentional" if the rater believed the clause implied an intentional action, performed by a living being with goals and intentions, and "non-intentional" if the rater believed the action could be used to describe an action made by an inanimate object. For example, "hide," "chase," and "talk" were coded as intentional and "move", "rotate", and "change direction" were coded as non-intentional. By this criterion, a minority of clauses (39%) were coded as intentional ($X^2 = 50.69$, $p < 0.001$). Forward videos elicited more intentional descriptions than backward, ($F(1, 10) = 10.41$, $p < .05$) suggesting that the videos were more interpretable forward than backward. As shown in Figure 3, coarse segmentation instructions produced a higher proportion of intentional verbs (*M* = .49, *SEM* = .05) than fine segmentation instructions (*M* = .39, *SEM* = .04), $F(1, 10) = 6.59$, $p < 0.05$. This finding supports the idea that coarse units are more conceptually determined than fine units.

Movements Clauses were coded for whether the verb strongly implied a motion vector component, for example, "move," "spin," "chase". The majority of descriptions reported physical movement (93%) ($X^2 = 797.92$, $p < 0.001$). Video direction did not affect the proportion of clauses reporting physical motion. As evident from Figure 3, fine segmentation instructions led to a higher proportion of motion verbs (*M* = .95, *SEM* = .02) than coarse segmentation instructions (*M* = .87, *SEM* = .04), $F(1, 10) = 6.29$, $p < 0.05$. Together with greater intentionality for coarse units, this result suggests that fine segmentation is more perceptually determined than is coarse segmentation.

Number of Actors Clauses were coded for the number of things doing the action (one or more than one). Participants were more likely to describe events involving a single agent than multiple agents ($X^2 = 297.07$, $p < 0.001$) but, coarse segmentation instructions (*M* = .31, *SEM* = .03) generally elicited a higher proportion of descriptions involving multiple agents than fine instructions (*M* = .23, *SEM* = .03), $F(1, 10) = 6.42$, $p < 0.05$ (See Figure 3).

Discussion

Participants segmented ambiguous events into the largest and smallest units that made sense. Half described what happened in each segment as they segmented. In previous work on everyday events using these methods, the vast majority of segments were described as actions on objects and segmentation was hierarchical, with the higher level parsed by objects or object parts and the lower level by refined actions on objects (Zacks, et al, 2001). Describing while segmenting the everyday events yielded greater hierarchical organization than segmenting alone.

The present study differs from previous work in that events used by previous researchers were relatively familiar, tightly organized sequences of actions performed by a single actor, primarily with hands, such as doing the dishes and fertilizing a plant. The present events were abstract, ambiguous animations of motion paths by geometric "trios," based on chase and hide-and-seek scenarios.

Despite these differences, there was a robust hierarchical alignment effect for the abstract events; that is, coarse unit boundaries coincided with fine unit boundaries more often than expected by chance. However, describing what happened in each segment did not increase hierarchical organization. Nor were the events described as actions on objects. Rather than being described as intentional actions, the events were described primarily as sequences of motion paths. For these events, then, segmentation was primarily based on bottom-up information, specifically, changes in motion paths, rather than on top-down information, such as goal-directed actions on objects.

Physical path changes are not unidirectional, in contrast to goal-directed actions. In the present study, hierarchical structure was perceived equally in forward and backward events, suggesting that understanding a goal structure is not necessary for hierarchical organization of events. Change in bottom-up physical information is sufficient to organize event perception. This provides empirical support for arguments that lower level physical changes can be used as a bridge to understanding goal structure, because attainment of goals is correlated with significant changes in physical action (Baldwin et al., 2001; Zacks & Tversky, 2001).

Our results are surprising given previous claims that the Heider-Simmel animation is spontaneously interpreted as intentional, socially-directed action. There are hints in the results that at least some of the actions are interpreted that way. Coarse-level units, which are more affected by top-down conceptual information, were more frequently described as intentional; they were also more frequently described as involving two actors. Forward videos elicited more intentional descriptions than backward videos.

Previous work has suggested (Newtson, 1973; Vallacher & Wegner, 1987) that when events are unfamiliar, they are interpreted at a finer level—a level of physical changes. If so, more experience with the videos should elicit more intentional interpretations. If segments are seen as intentional, will the same segments be identified after repeated exposure as initially?

Exp 2: Abstract Events Viewed Repeatedly

Method

The method was identical to that used in Experiment 1, except that prior to receiving segmentation instructions, participants viewed each video 5 times consecutively (in the order that the films would later be segmented). They were told to pay careful attention to the videos as they would be later asked to describe what happened. After 5 viewings, participants were asked to write a brief description of each video. They were not told how to interpret the videos. Thirty-two Stanford undergraduates completed the experiment in exchange for pay or extra credit.

Results and Discussion

Hierarchical Structure The mean number of fine breakpoints (M = 15.30, SEM = .69) was significantly greater than the mean number of coarse breakpoints (M = 5.77, SEM = .34), $t(63) = 16.02$, $p < 0.001$, indicating that participants were segmenting the films according to the instructions. As shown in Figure 1, there were reliably fewer breakpoints identified in Experiment 2 than in Experiment 1 ($F(1, 76) = 12.78$, $p < 0.01$), though the difference between the mean number of coarse and fine breakpoints appears to be consistent across experiments. The fact that participants in Experiment 2 identified fewer breakpoints overall is consistent with the idea that familiarity with an activity leads to a coarser level of segmentation.

As in Experiment 1, a reliable hierarchical alignment effect was found, $t(63) = 2.87$, $p < 0.001$. Figure 2 shows the observed and expected (Null) average distances for Experiments 1 and 2. Though the observed and expected distances are higher for Experiment 2, likely due to the lower frequencies of coarse and fine breakpoints, the difference between the observed and expected distances did not differ between the two experiments, $F(1, 76) = .06$, p = 0.81. As in Experiment 1, there was no effect of video direction, $F(1, 16) = .069$, $p = .80$, nor of describing $F(1, 16) = .897$, $p = 0.36$.

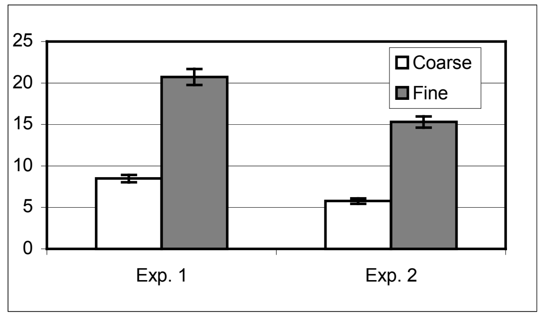

Figure 1: Mean number of breakpoints as a function of segmentation level (Coarse, Fine) for Experiments 1 and 2.

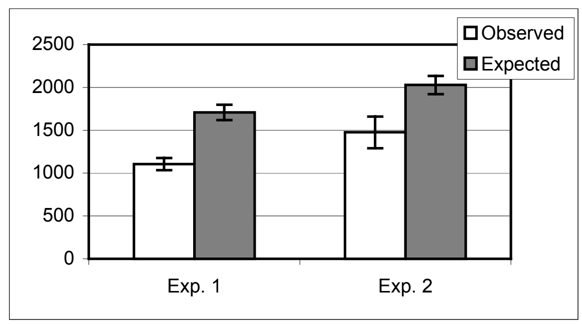

Figure 2. Observed and expected average distances between fine and coarse breakpoints for Experiments 1 and 2.

Verbal Descriptions A total of 789 clauses from 16 participants were coded as in Experiment 1. Inter-rater reliability was above 0.92 for all categories. In contrast to Experiment 1, most of the clauses (69%) were coded as intentional, goal-directed actions ($X^2 = 121.63$, $p < 0.001$) suggesting that repeated viewings of the film successfully improved their interpretability. Consistent with Experiment 1, most of the descriptions involved physical movement ($X^2 = 317.20$, $p < 0.001$) and a single agent rather than multiple agents ($X^2 = 202.80$, $p < 0.001$). Forward videos again produced reliably more clauses suggesting intentionality than backward videos, $F(1, 8) = 10.61$, $p < 0.05$.

Figure 3 provides a comparison of the results from Experiments 1 and 2. As in Experiment 1, coarse units were more intentional than fine units, but this effect was only marginally reliable, $F(1, 8) = 4.21$, $p = .07$. Fine units were more likely to involve physical motion than coarse units, $F(1, 8) = 10.88$, $p < 0.05$. In contrast to Experiment 1, there was no reliable effect of segmentation level on the proportion of clauses that referred to multiple actors, $F(1, 8) = 2.03$, $p = 0.19$.

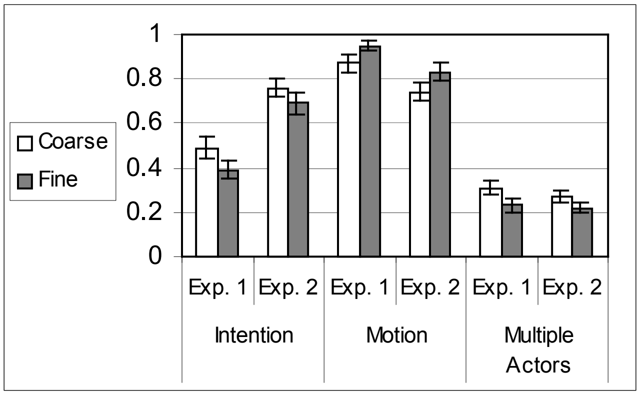

Figure 3. Proportion of actions describing intention, motion, and multiple actors as a function of segmentation level (Coarse, Fine) for Experiments 1 and 2.

Breakpoint Agreement To determine whether the breakpoints selected in Experiment 1 differed from the breakpoints selected in Experiment 2, two analyses were performed: First, for Experiments 1 and 2, each film was divided into 85, 1-sec bins and the proportion of participants who had selected a given bin as a breakpoint was determined for each bin. Pearson correlations were then run for each of the four videos and the two segmentation levels, comparing the proportions from the 85 bins in Experiment 1 to Experiment 2. As Table 1 shows, reliable correlations were found for each film. This implies strong agreement on breakpoint selection, despite differences in how the segments were interpreted. Figure 4 shows one graphical example of the similarity between breakpoint distributions in the two experiments.

Table 1: Results of correlation analysis of bin proportions in Experiments 1 and 2; *$p < 0.01$.

	Fine	Coarse
"Chase"		
Forward	.536*	.390*
Backward	.627*	.633*
"Hide and Seek"		
Forward	.553*	.560*
Backward	.451*	.398*

Second, we divided each film into 17, 5-sec bins and performed a series of Chi-Square tests to see if the frequency of key presses in the bins differed between Experiments 1 and 2. For all conditions, the Chi-Square test revealed no relationship between the frequency of breakpoints in each bin and Experiment. The X^2 values ranged from 12.18 to 16.68, with p values ranging from 0.41 to 0.73. Thus, the two groups of observers selected approximately the same breakpoints with the same frequency, regardless of their level of experience with the films.

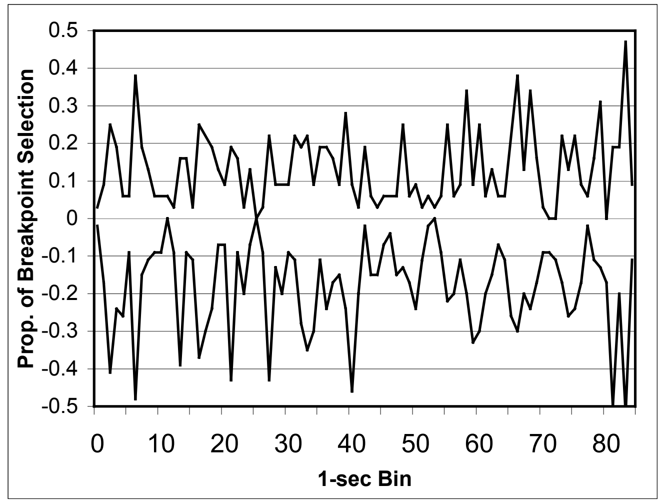

Figure 4. Proportion of times that a 1-sec bin for *chase* (forward) was selected as a fine or coarse breakpoint. Proportions for Exp. 1 (View Once) are shown as *negative* to facilitate visual comparison to Exp 2 (View 5 Times).

General Discussion

During an initial viewing, observers were able to segment unfamiliar, ambiguous, and abstract events hierarchically; however, they interpreted these events in terms of physical motions rather than intentional actions. With repeated exposure, observers came to interpret the events as sequences of intentional actions.

The changes in interpretation with exposure were not accompanied by changes in segmentation. With and without exposure, the events were segmented hierarchically, and the same segments were discriminated. Fewer segments were identified following exposure, consonant with more conceptual interpretations: intentions group action segments into larger wholes by conferring relations on them. This is reminiscent of the Gibsons' observations (Gibson & Gibson, 1955) on perception. Like wine-tasters, observers of events come to perceive more in what is out there (perceiving intentionality, however, was not what the Gibsons meant).

With and without exposure, the coarse level units were interpreted more conceptually than the fine level units. Coarse level descriptions more frequently referred to intentions and to more than one agent; fine level descriptions more frequently referred to motion paths. In contrast to segmenting actions by hands, segmenting actions by "feet", that is, changes in motion paths, was not more hierarchical when the actions were described than when they were only segmented. Moreover, forward and backward versions were equally hierarchical, even though forward versions were interpreted more intentionally, suggesting that for these events, salient changes in motion paths underlie segmentation.

Importantly, intentions and motion are more tightly coupled in motion paths than in hand manipulations and body movements, offering a compelling explanation for our findings of consistent segmentation in the face of interpretation change. The changes in motion paths were primarily changes in direction, though some were changes in manner and speed of motion. Hand manipulations and body actions in events such as making a bed or fertilizing a plant are far more subtle and variable, with some more closely linked to goals and others incidental. In some cases, incidental or less central hand and body motions may be more salient than those closely linked to goals; for example, bending over to pick up a blanket may be more salient than tucking in a corner and positioning a plant may be more salient than opening a fertilizer container. This explains why Zacks et al. (2001) found small effects of activity familiarity on hierarchical perception for these kinds of events: goal-based event schemas should allow discrimination between goal-related and incidental body motions.

The correlation between salient motion changes and intentions at event boundaries suggests how event segmentation can occur in the absence of interpretation and in fact, facilitate construction of interpretations. Baldwin et al. (2001) have argued that parsing events *leads* to intentional knowledge, rather than resulting from it. Bolstering this hypothesis, infants with only limited, if any, knowledge of functions and intentions of actions, nevertheless parse the behavior stream in the same way adults do (e. g., Baldwin et al., 2001; Sharon & Wynn, 1998).

This analysis can explain how goal-related schemas for highly familiar activities—such as those used by Zacks et al. (2001)—are formed. The present studies suggest that a hierarchical schema for goals and functions of low-level events might begin, rather than end, with hierarchical perception of motion events based on salient physical changes in the world. Once hierarchical event schemas are formed, they can be used to anticipate event segments, yielding stronger alignment effects for more highly familiar events, at least those performed primarily by hands, as Zacks, et al. (2001) found. Describing events can augment this process, by calling attention to the intentional, goal-related aspects of events, again consistent with the findings of Zacks, et al. (2001). What counts as salient physical changes seems to depend in part on whether the actions are performed primarily by hands or by feet. Actions by hands are segmented by objects (Woodward, 1998; Zacks, et al, 2001). Actions by feet are segmented by changes in motion path, primarily changes of direction.

Acknowledgments

Conversations with Helen Hwang, David Lang, Jeff Zacks, Catherine Hanson, and Stephen Hanson have been stimulating and helpful in the development of our ideas. We are grateful to them, and to the following grants: Office of Naval Research, Grants Number NOOO14-PP-1-O649, N000140110717, and N000140210534 to Stanford University.

References

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*. Mahwah, NJ: Lawrence Erlbaum Associates.

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*. New York, NY: Appleton-Century-Crofts.

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*. New York, NY: Appleton-Century-Crofts.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32-41.

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

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28,* 28-38.

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

Rosch, E. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Sharon, T., & Wynn, K. (1998). Individuation of actions from continuous motion. *Psychological Science, 5,* 357-362.

Tverksy, 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.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.

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

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