Perspective distortion of trajectory forms and perceptual constancy in visual event identification

EMILY A. WICKELGREN
California State University, Sacramento, California

and

GEOFFREY P. BINGHAM
Indiana University, Bloomington, Indiana

Previous studies have shown that people can use the information in trajectory forms to recognize visual events. A trajectory form is composed of the path of motion and the change in speed along that path. In past studies, however, only sensitivity to trajectory forms viewed from a single perspective was examined. The optical components change when an event is viewed from different perspectives, and the projected form of the trajectory is transformed. Does event recognition exhibit constancy despite these changes? In Experiment 1, participants were familiarized with five different trajectory forms viewed from a single perspective. Then the participants had to identify the same events viewed from different perspectives: from the side, at an angle, and entirely in depth. The participants exhibited perceptual constancy. Experiment 2 revealed, however, that both the change in optical components and the perspective transformations affected recognition.

Imagine yourself at a baseball game. There are children running up and down the aisles, confetti blowing across the field, a vendor pouring you a drink, a baseball bouncing along the infield, and a player catching the ball and throwing it to the pitcher. You would be able to identify these visual events with ease: rigid events such as the bouncing ball, nonrigid events such as the confetti or spilled drink, and animate events such as the children running and the player throwing. The ability to recognize and identify different types of events is as apparently facile and as functionally relevant as our ability to recognize objects. Although the appearance of unmoving surfaces and objects in events certainly contribute to our perceptions, it has been demonstrated that adults are extremely sensitive to motion in events and are able to recognize events by using only the motion information (Johansson, 1973, 1976). In patch-light displays, events are filmed with luminescent patches on objects so that only the bright patches are visible against a dark background. This technique isolates motion as the only source of information about an event. Static patch-light images contain insufficient structure to allow observers to recognize an event, but when the images flow over time, the objects and events become easily recognized. Bingham, Rosenblum, and Schmidt (1995) showed observers patch-light displays of both a freely swinging pendulum and a pendulum that was being moved by hand. Only the patches on the pendulum were visible. In the two cases, the motions had the same path of movement and the same frequency, but the pattern of the change in velocity during the swing was different. Adults could discriminate between the events and were often able to identify the events as well. Indeed, the ability of adults to recognize events was demonstrated with a variety of inanimate events, including bouncing balls, splashing water, and wind-blown leaves. Johansson (1973, 1976) had shown previously that observers are able to identify complex animate events. When shown patch-light events of people walking, running, or dancing, observers easily identified the events. In fact, sensitivity to the information in motion is so good that adults often can identify friends when shown only the motions generated during their friends' walking (Cutting & Kozlowski, 1977; Frykholm, 1983).

Many researchers have discussed the type of information that people might use to identify events. Johansson (1950) argued that observers use velocity vector analysis to identify events, subtracting out the relative and common motion components in an event. This led to the development of structure-from-motion studies, in which motion-generated information for object recognition is investigated. Following the discovery that observers can detect fatigue and amounts of lifted weight from human motion in patch-light displays, Runeson and Frykholm

(1983) proposed the kinematic specification of dynamics principle (see also Runeson, 1977, 1977/1983). The principle states that the underlying dynamical constraints on an event (gravity, inertia, etc.) determine specific kinematic (i.e., motion) patterns and that observers can determine the nature of an event from the resulting patterns of motion. The specific nature of the patterns that allowed recognition of the human motions was not described, although a potential source of information about relative mass in collisions was described and investigated (Runeson, 1977, 1977/1983; Runeson & Vedeler, 1993), with some ensuing controversy (Gilden & Proffitt, 1989, 1994; Runeson, 1995).[1]

Following an investigation of trajectories as information about lifted weight (Bingham, 1987b), Bingham (1995) proposed that trajectory forms are the information used to identify events. Just as in object recognition, it is form that is preserved over changes in viewing distance, and thus, the forms of events must be used to recognize events. A trajectory form consists of both the three-dimensional shape of the path of motion and the shape of the speed profile along that path. A trajectory form is generated by continuous change in velocity (i.e., speed and direction) over the course of an event. Trajectory forms can be plotted in a phase (or state) space with position and speed axes (see Bingham, 1995, for examples, as well as the figures referenced in the Method sections of this article). Although a full representation would be a plot of trajectories in an event space with position, speed, and time axes, structure-from-motion studies have shown that observers are sensitive primarily to the spatial derivatives of velocity (which would be used to describe the shape of a phase portrait), rather than to the temporal derivatives (which would be used to describe the shape of the velocity time series; e.g., Perotti, Todd, Lappin, & Phillips, 1998).[2] In addition, space is the natural frame of reference in which the trajectory of an event is seen to evolve.

Muchisky and Bingham (2002) investigated the sensitivity of adults to the speed component of trajectory forms by systematically manipulating the shape of the speed profile of an oscillator in one of four ways while holding the path and duration constant. The computer-generated event consisted of a circle oscillating back and forth along a straight horizontal path on the computer screen. The circle's speed along that path varied as a function of the nonlinear oscillator used to generate the motion. The shape of the speed profile was isolated by varying the spatial scale of events compared in a two-alternative forced-choice task. Muchisky and Bingham found that adults could discriminate between the events up to small differences in form, showing that adults are sensitive to variations in the shapes of speed profiles. Wickelgren and Bingham (2001) have also shown that infants as young as 8 months are sensitive to trajectory forms.

People rarely see a given event from the same perspective each time they see it. Think of the baseball game again. When the pitcher throws the ball, that same event is being viewed by people from many different perspectives. The first baseman will have a view of the pitch from the side, while the catcher will see the ball coming directly toward him. Spectators in the stands will see the pitch from various oblique angles, depending on where they are sitting. Muchisky and Bingham (2002) showed that adults can use trajectory forms to identify simple oscillating events seen from the side, but they did not investigate other perspectives. We now will investigate this issue of perceptual constancy in event identification. The question is whether adults can distinguish and identify events that vary only in their speed profile when the events are viewed from different perspectives (e.g., end-on as opposed to from the side).

Two changes occur in visual information when events are viewed in depth, as opposed to from the side. First, with changes in perspective, the same forms of motion will be projected into very different optical components. When straight path events are seen from the side, the information about the form of the event is variation in the rate of image translation. When an event is viewed from a perspective orthogonal to the side view, so the event is seen end-on in depth, the form of the event is reflected entirely in variations of the optical expansion rate. When an event is viewed from a perspective in between the side and the depth perspectives (i.e., viewing from an oblique angle), the information about the event is expressed by both image translation and image expansion.

The second change in visual information, related to such changes in perspective, is that when the form of an event is viewed in depth or at an angle, perspective transformations occur in the form. The type of transformation is illustrated by the increase in the optical expansion rate that occurs as an object approaches an observer at a constant velocity (called *looming*). This phenomenon is related to motion parallax, in which an object moving at a given speed far from an observer has a slower optical velocity than does an object moving at exactly the same speed but at a distance closer to the observer. Simply, optical velocity is inversely proportional to viewing distance. This transformation could potentially change the perceived identity of events or diminish the distinctiveness of alternative forms of motion, making it difficult to tell the forms apart (see Experiment 1 for a description of trajectory forms and their changes). This transformation is exactly the same as the perspective transformations that occur when objects are viewed from different orientations.

Research in the field of object recognition has shown that people exhibit perceptual constancy despite changes in the perspective on an object. Observers are able to recognize the same object when it is shown at different orientations with respect to the observer, although reaction times are often longer with a greater change in perspective (e.g., Shepard & Metzler, 1971). Other studies on object recognition have also shown evidence for viewpoint specificity (Newell, Ernst, Tjan, & Bülthoff, 2001; Tarr, Pepper, Hayward, & Gauthier, 1998). Bingham (1995) proposed that events are spatiotemporal objects.

He proposed that events have form and structure just as objects do and that, just as object form is a basis for object recognition, trajectory form is a basis for event recognition. If this is so, event recognition should exhibit constancy despite changes that accompany changes in perspective, just as does object recognition. Perceived identity should be preserved despite both changes in the components of optical information (image translation vs. image expansion/contraction) and perspective transformations.

In the following experiments, we investigated the issue of perceptual constancy in event recognition by testing whether adults could identify the same trajectory forms viewed from different perspectives. To do this, we isolated trajectory forms as information for study. Typically, in naturally occurring events, trajectory forms would be combined with relative motions of the sort studied in structure-from-motion research, as has been discussed in Bingham (1995). So, for instance, a ball rolling in a gutter would exhibit a harmonic (oscillatory) trajectory in an optic flow also containing structure-from-motion information generated by the ball rotation. We isolated trajectory forms by projecting only the contour of a moving ball into an optic flow display. The particular trajectory forms used in these experiments were inspired by previous studies in which the recognition of animate and inanimate events was investigated (Bingham et al., 1995).

EXPERIMENT 1
Identification of Speed Profiles in Depth

In Experiment 1, participants were familiarized with five different trajectory forms viewed from a single perspective. Then the participants tried to identify the same events viewed from different perspectives: from the side, at an angle, and entirely in depth.

Method

Participants. Sixteen adults participated in this experiment. Seven observers were paid $6 per hour; the other 9 observers received course credit. The observers were told that they would receive a $20 bonus if they had one of the two highest accuracies in the study, as a motivation to try their best.

Display generation. The simulated events consisted of the oscillatory motion of a ball along a straight path. Oscillation was at 0.5 Hz, and the ball diameter was varied between 1.5 and 2.5 cm. There were three viewing conditions in the experiment: side view, oblique view, and depth view. In the side view, each display consisted of a circle that translated back and forth across the screen along a straight path. The form of the events was isolated by varying the displacement from trial to trial (9–12 cm), while holding the period (2 sec) constant across trials. These variations prevented the observers from using any artifactual information, such as peak velocity, to identify the different events. The observers were seated approximately 50 cm from the monitor, and the side-view events were simulated to be at the distance of the surface of the computer monitor. The range of visual angle for the displays was between 10° and 13°, for an average optical velocity of approximately 11° per second. This and the other optical velocities used in this study were within the range at which differential thresholds were found to be at a minimum (Orban, de Wolf, & Maes, 1984).

In the depth view, each display consisted of a circle in the center of the screen that changed its diameter. The image size of the circle was determined by the simulated distance of the ball from the observer and the size of the ball. Thus, the shape of the event's speed profile was projected purely into the change in image size of the circle. The amplitudes of the oscillations in depth varied between 72 and 96 cm. The increase in amplitude was necessary to produce changes in image size on the screen that were well above threshold. Since speed profiles are invariant across changes in amplitude, this would not change the trajectory form. The point on the path closest to the observer was simulated to be at the distance of the surface of the computer monitor. At the farthest point from the observer, the image size of the ball was approximately 1° of visual angle, and at the closest point it was about 3° of visual angle. The average optical expansion or contraction rate was, therefore, approximately 4° per second.

In the oblique view, the circle translated back and forth on the screen, while its diameter changed according to the simulated ball depth. The screen amplitude of the event as the circle moved back and forth was matched to the screen amplitude in the side-view condition (9–12 cm), which partially determined the simulated world amplitude of the oblique-view event. The simulated amplitude of the event was matched to the simulated amplitude of the depth-view event as well. Together, these determined the degree of the path's tilt away from the observer's line of sight (Figure 1). The range of the viewing angles was 10°–13°, because the screen amplitude was equated to that in the side-view condition. The optical velocity of the oblique-view event was, therefore, approximately 11° per second. The range of simulated world amplitudes for the oblique view was 76–102 cm.

Five different speed profiles were generated: asymmetric, bimodal, constant, harmonic, and stretched (Figure 2). The asymmetric speed profile was generated using a van der Pol oscillator (Equation 1), where λ was a positive coefficient on the nonlinear damping term. As the value of λ was increased, the peak velocity of the event increasingly departed from the midpoint of the path, yielding the characteristic asymmetry. For each of the displays, we chose coefficients that were above the 75% threshold for discrimination found by Muchisky and Bingham (2002). For the asymmetric speed profile, $\lambda = 1.75$, and $k = 1.9$.

$$\ddot{x} = -\lambda\, \dot{x}\left(x^2 - 1\right) - kx. \qquad (1)$$

The bimodal speed profile was generated using the equation for a hard spring (Duffing oscillator), where c is a constant controlling the period of the event and k is the spring term (Equation 2). For the bimodal speed profile, $c = 1$, and $k = 2.5$.

$$\ddot{x} = -c\left[x - k\left(x^3\right)\right]. \qquad (2)$$

Increasing k causes the speed profile to flatten and, eventually, to dip in the middle to form a bimodal profile with two peaks that occur near each end of the path. In the constant speed profile, the object was constant in speed until, at each endpoint, the circle suddenly and discontinuously reversed direction. The harmonic speed profile was created using a harmonic oscillator (Equation 3), where k is a linear stiffness term that was set to 1.58.

$$\ddot{x} = -k\, x \qquad (3)$$

The final speed profile, stretched, was produced using the equation for a soft spring (also a Duffing oscillator; Equation 4).

$$\ddot{x} = -c\left[x + k\left(x^3\right)\right]. \qquad (4)$$

As with the equation for a hard spring, the constant c controls the period of the event. As the spring term, k, increases, the speed profile becomes increasingly peaked or stretched. For the stretched event, $c = 7.52$, and $k = .6$.

These trajectory forms relate approximately to naturally occurring events as follows. Harmonic motion is typical of inanimate oscillatory events, such as wind-blown tree limbs and other pendulum
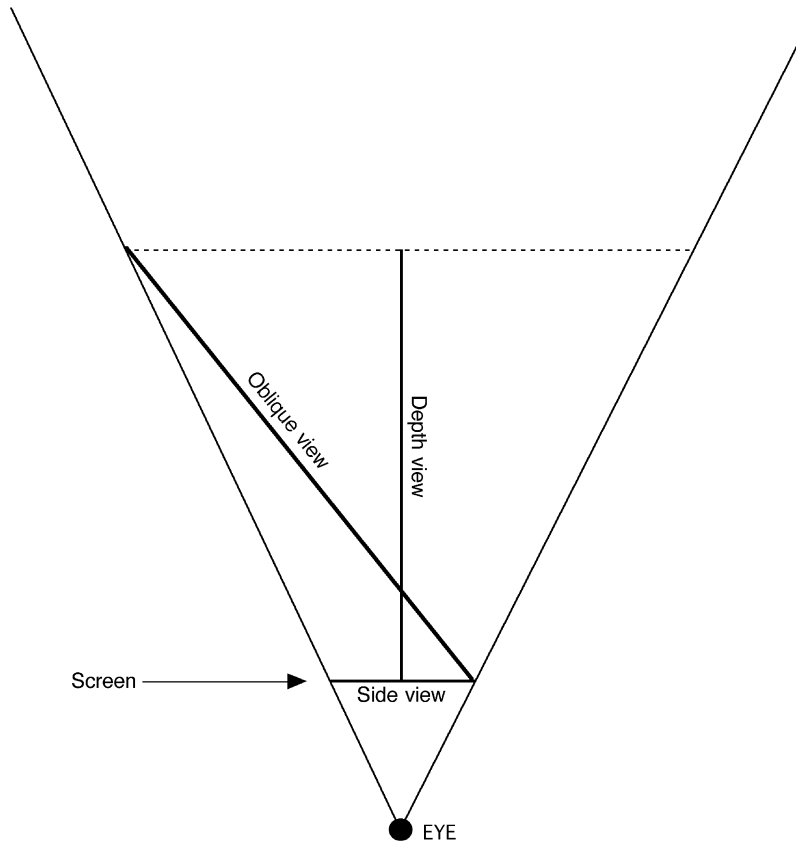
Figure 1. Figure depicting how the oblique-view event was created. The amplitude of the depth-view event and the amplitude of the side-view event on the screen determined the actual amplitude of the oblique-view event.

motions. The asymmetric, stretched, and bimodal motions represent forms that can be exhibited by animate limb movements. The van der Pol and Duffing equations used to generate these motions have also been used to model human limb movements (e.g., Kay, Kelso, Saltzman, & Schöner, 1987). The constant velocity motion represents hard-geared motions of mechanical devices or toys.

Each of these equations specifies the event in the side view. When these same events were projected into depth, however, the profiles were subjected to the characteristic transformations described earlier. Figure 3 shows the effect of the perspective transformation on the side-view speed profiles. In the oblique view, both the image size and the position on the screen were partially determined by distance from the observer. Therefore, similar perspective distortions obtained for both the horizontal (image translation) and the depth (image expansion/contraction) components (Figure 4).

Design and Procedure. The experiment consisted of two phases, a training phase and a testing phase. The purpose of the training phase was to ensure that the observers could reliably identify which speed profile they were viewing in the side view, before testing for perceptual constancy in the oblique view and the depth view. At the beginning of the training phase, the observers were shown examples of each type of event in the side view to familiarize them with each speed profile. They were then presented with a block of 25 trials (five events each seen five times in the side view). For each trial, the observers viewed a display and then had to identify which speed profile they had seen (A, B, C, D, or E). Responses were made using the mouse and clicking on a button in a dialog box. After each judgment, the observers were given feedback about their response. The

observers were required to achieve an accuracy of 80% for each speed profile before advancing to the testing phase. If they did not achieve this criterion, they repeated familiarization and then another block of 25 trials. Once they had reached the 80% criterion, they began the testing phase. At the beginning of the testing phase, the observers were instructed that their task would be the same as before: to identify which event they had just seen. They were told that the events would now not always be traveling side to side but might be coming toward and away from them in depth or at an angle toward and away from them in depth. The testing phase consisted of six blocks of 15 trials each (five speed profiles each randomly presented in one of the three orientations). During the testing phase, the observers were not given feedback about their performance.

Results and Discussion

The first dependent variable calculated was accuracy (percentage correct).[3] It was calculated for each orientation in the test phase and was used to determine whether the observers were able to identify specific trajectory forms in different orientations. The mean percentages correct (and standard deviations) for the side-view, oblique-view, and depth-view cases were 88.75% (8.85), 53.54% (10.36), and 45.21% (10.18), respectively, where chance performance was 20% correct. The observers were performing above chance in each viewing orientation, indicating that, in general, they could recognize the events in the oblique
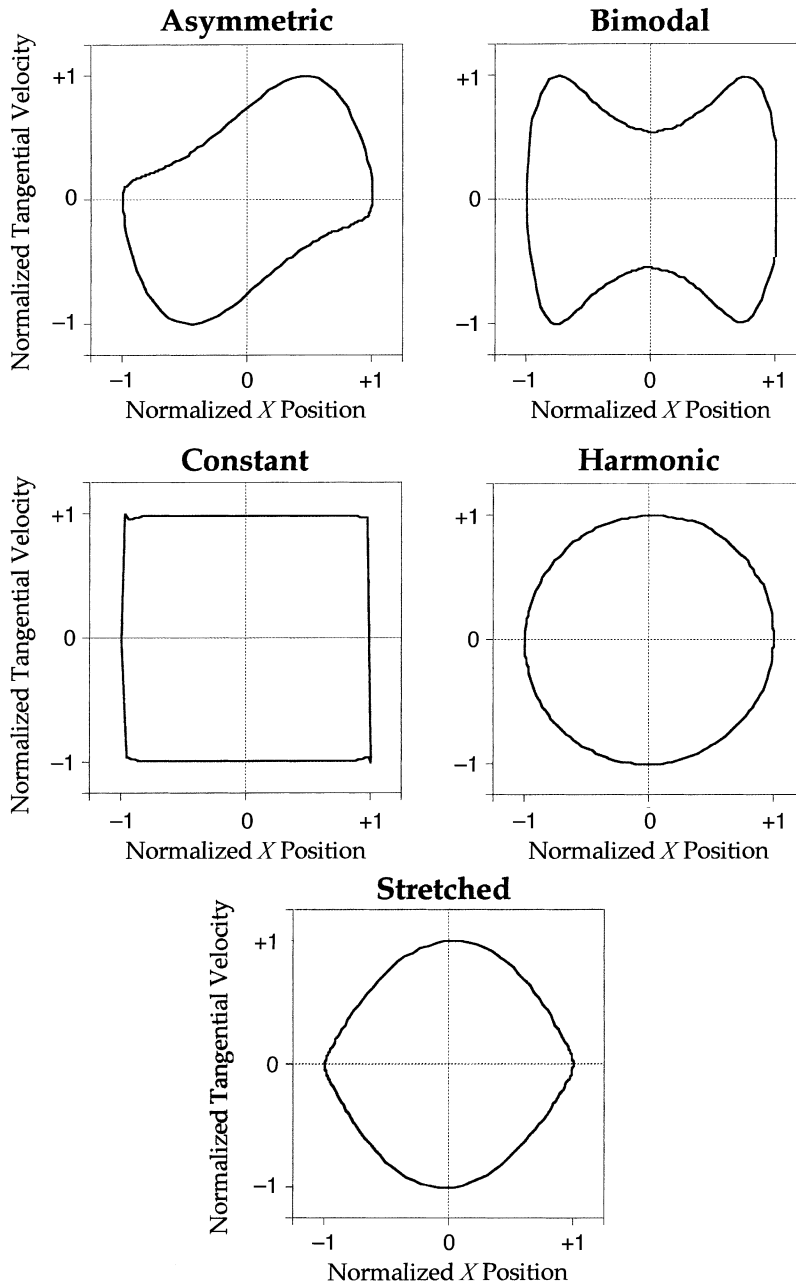
Figure 2. The five side-view speed profiles used in Experiment 1. Speed is plotted as a function of position for each speed profile. The units have been normalized.

and depth conditions as the same events as those they had originally seen in the side-view condition. Because the data were frequencies, a logit($\hat{p}$) transform was applied to yield a normal distribution (Johnson & Wichern, 1998). A repeated measures analysis of variance (ANOVA) revealed that performances in the three conditions were significantly different [$F(2,15) = 11.66, p < .01$]. Pairwise comparisons showed that the observers were less accurate for the oblique view than for the side view [$t(15) = 3.25, p < .01$] and even less accurate for the depth view

than for the oblique view [$t(15) = 2.84, p < .05$]. We would expect a high percentage correct for the side-view condition, because the participants were required to achieve at least 80% correct in the training phase before moving on to the testing phase. However, it was unclear whether the participants would be able to identify the speed profiles when they were viewed in depth, because the information was contained in different optical components and because of the optical distortions described earlier (Figure 3). The participants were above chance in
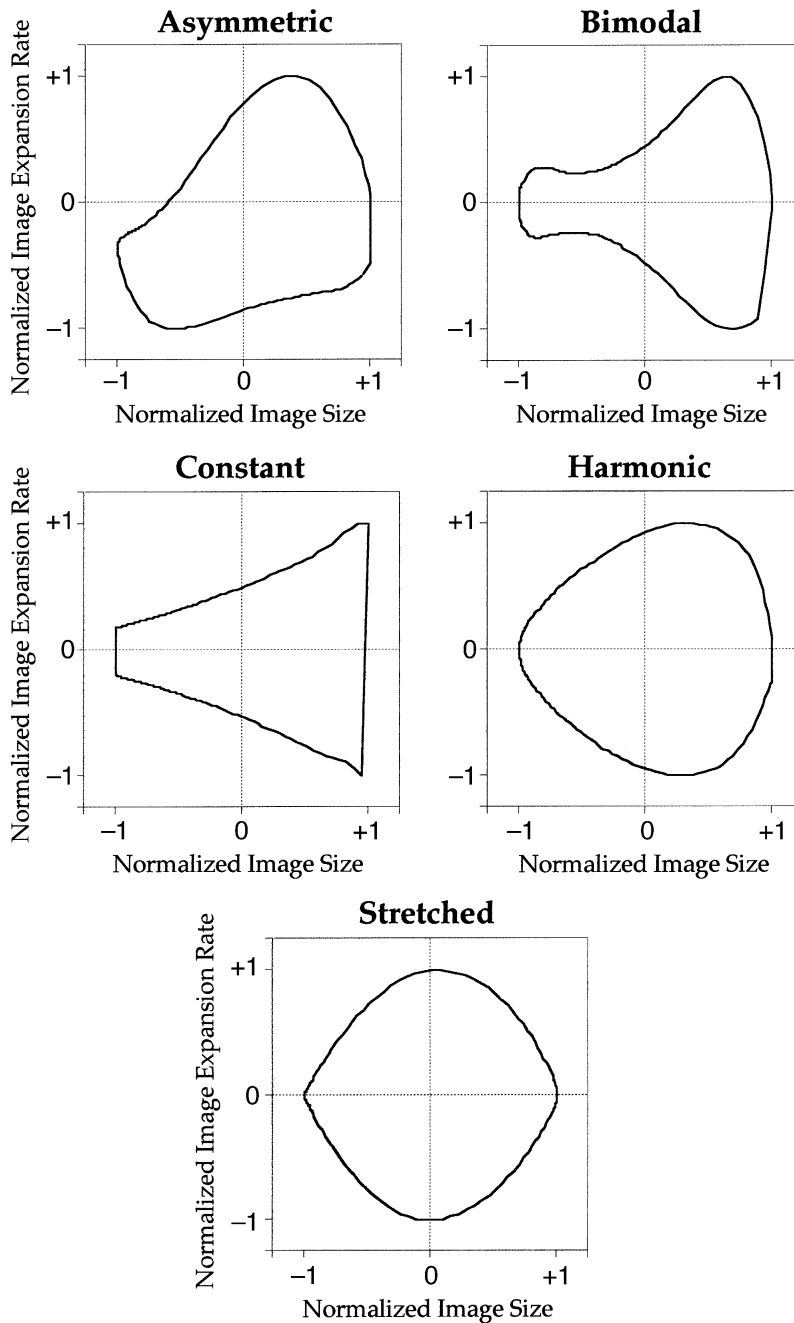
Figure 3. The depth-view analogues to the original five speed profiles used in the side-view condition in Experiment 1. The depth-view profiles are a result of the perspective distortions. Image expansion rate is plotted as a function of image size for each speed profile. The units have been normalized.

their identifications for all three viewing orientations. So they exhibited perceptual constancy for speed profiles.

Nevertheless, it was clear from the drop in accuracy for the oblique- and the depth-view conditions that the observers found it more difficult to identify the displays in those conditions. To determine how well the observers discriminated each display, we computed a d' for each

individual display against the combination of the other four displays in each viewing orientation (Table 1). For example, in the comparison of the asymmetric display with the others combined, the hit rate was the number of times the observers chose *asymmetric* when they were shown the asymmetric display, the false alarm rate was the number of times they chose *asymmetric* when they
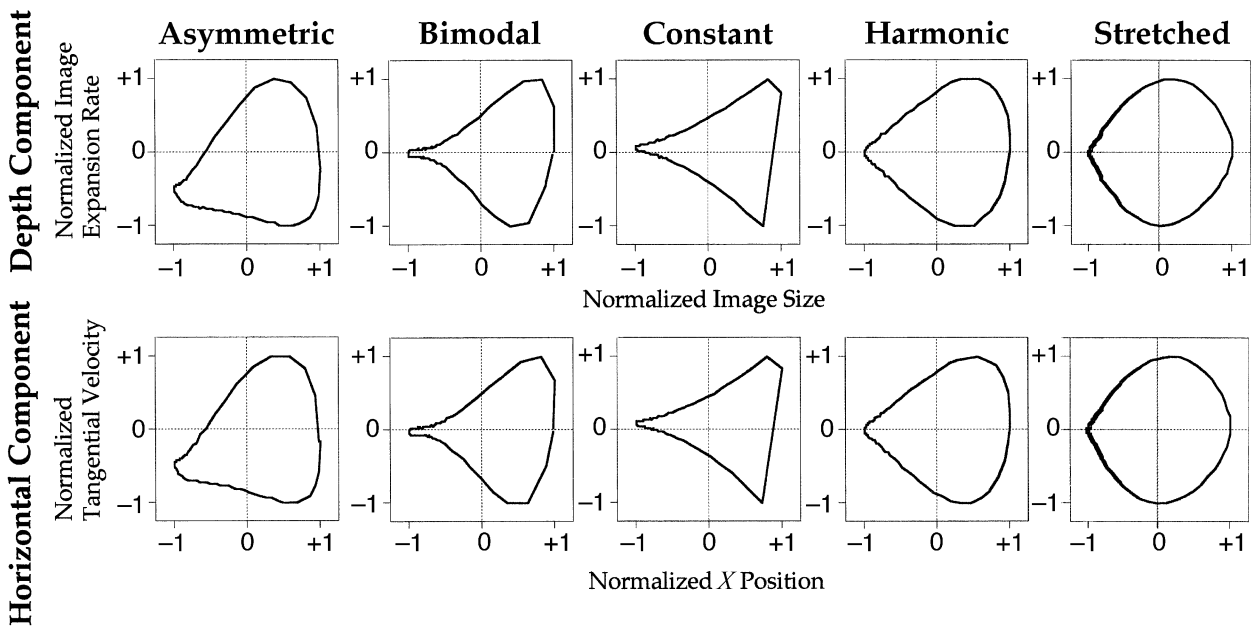
Figure 4. The oblique-view analogues to the original five speed profiles used in the side-view condition in Experiment 1. Plots for both the image expansion/contraction component and the translation component are shown for each event. The units have been normalized.

were, in fact, shown one of the other four displays, the miss rate was the number of times they chose another display when they were shown the asymmetric display, and the correct rejection rate was the number of times they chose another display when they were shown another display. The rule of thumb, provided by Macmillan and Creelman (1991), is that a d' above 1.0 indicates sensitivity to the stimuli and, in this case, perceptual constancy, whereas anything below 1.0 indicates that the two stimuli were confused and there was no perceptual constancy for that event. The d' for the asymmetric event in the side view was 2.89, which is well above 1.0, indicating that the observers had no difficulty discriminating the asymmetric speed profile from the other speed profiles. Overall, in the side-view and oblique-view conditions, the participants did not appear to have trouble identifying the displays (all d's were above 1.0). However, in the depth-view condition, sensitivity dropped below 1.0 for the bimodal and stretched displays ($d' =$

0.68 [29% correct] and $d' = 0.77$ [34% correct], respectively). So the observers were no longer able to identify these two displays when they were viewed in depth. However, this analysis did not indicate the particular events that were confused. So we also computed a d' for each pairwise comparison, comparing each display with every other display in each viewing condition (asymmetric vs. bimodal, asymmetric vs. constant, asymmetric vs. harmonic, etc.; Table 2).

The participants did not have any problems discriminating between speed profiles in the side-view condition (all d's > 2.50). However, in the oblique view, sensitivity began to drop. Note that it is with the oblique view that we first see the perspective distortions in both the horizontal component and the image expansion component. Although sensitivity dropped, the participants discriminated all speed profiles except the harmonic and the stretched displays, where d' dropped to 0.91. An even bigger drop in sensitivity occurred in the depth-view condition, where the only information available was from changes in image size, also containing the perspective distortions. The results showed that the participants were still able to discriminate between all the pairs of displays except two. Consistent with the results for the oblique view, the participants had difficulty distinguishing the harmonic and the stretched speed profiles ($d' = 0.66$). Now, in addition, the participants were no longer able to distinguish the bimodal and the constant speed displays ($d' = 0.83$).

As can be seen in Figure 3, the optical distortion changed the harmonic and the stretched displays so that they became more similar (both became more oval in shape).

Table 1
d' Calculated in Three Viewing Conditions for Each Event Versus the Remaining Four Combined in Experiment 1

Event Contrast	d'		
	Side View	Oblique View	Depth View
Asymmetric vs. others	2.89	1.32	1.14
Bimodal vs. others	3.49	1.26	0.68
Constant vs. others	3.59	1.71	1.22
Harmonic vs. others	3.26	1.23	1.09
Stretched vs. others	2.86	1.06	0.77
Mean sensitivity	3.22	1.32	0.98
Median sensitivity	3.26	1.26	1.09

		d'	
Contrasted Pair	Side View	Oblique View	Depth View
Asymmetric/bimodal	3.71	1.49	1.11
Asymmetric/constant	4.81	2.51	1.85
Asymmetric/harmonic	4.79	2.71	2.31
Asymmetric/stretched	2.62	1.48	1.20
Bimodal/constant	3.75	1.83	0.83
Bimodal/harmonic	5.07	2.45	1.59
Bimodal/stretched	4.30	2.04	1.78
Constant/harmonic	3.68	2.16	1.95
Constant/stretched	5.01	2.67	2.31
Harmonic/stretched	3.76	0.91	0.66
Mean sensitivity	4.15	2.03	1.56
Median sensitivity	4.03	2.10	1.69

The same was true of bimodal and constant displays, both of which exhibited funnel-like shapes. Muchisky and Bingham (2002) showed that as two speed profiles become more similar in their shape, adult ability to discriminate them becomes poorer, just as one would expect. The two display pairs with the highest d' (and hence, the least confused) were the asymmetric and harmonic displays and the constant and stretched displays. As can be seen in Figure 3, each member of those pairs had a very different shape. The shape of the asymmetric speed profile was skewed, having a very rapid expansion in image size as the object approached but a very slow contraction of the image size as the object retreated. The change in image size for the harmonic display (as well as the other displays) was the same when the ball approached as when it retreated.

Overall accuracy was better in the oblique-view condition than in the depth condition. One might surmise that the addition of an image translation component to the expansion component could have been responsible for the improved performance. It is important to note, however, that the image translation component in the oblique-view condition was also affected by the perspective transformations (Figure 4). Nevertheless, if it is easier to use the information in image translation, that may explain why performance was slightly better in the oblique-view condition even though both components had distorted speed profiles.

EXPERIMENT 2
Effect of Perspective Distortion and Optical Components on Identification of Speed Profiles in Depth

Experiment 1 showed that observers had the greatest difficulty discriminating between the displays in a depth-view condition, whereas they were much better in a side-view condition. One possible reason for this difference in performance is that the observers were not trained with feedback on depth-view displays, so performance would

naturally be lower than in the side-view condition, where they did get feedback. We tested this in Experiment 2. The observers were given feedback in every viewing condition until their performance reached an asymptote. Another possible reason for the difference in performance could be that the depth-view displays were more similar in shape to one another and, thus, harder to discriminate than the side-view displays. This would be due to the perspective transformations. Indeed, there was also a drop in performance in the oblique-view condition, where the perspective distortions were also present. We tested this possibility as well in Experiment 2. The observers were shown displays with both distorted and undistorted shapes in both side-view and depth-view conditions. The change in shape without change in optical components tested the effect of perspective distortions on recognition. A final possibility is that the decrement in performance was due to the change in optical components (translation vs. expansion/contraction). If sensitivity to variations in expansion rates is less than sensitivity to optical translation velocities, performance should be poorest for the perspective with no translation component. This possibility is suggested by the fact that performance dropped again from the oblique- to the depth-view conditions. Finally, we tested this possibility in Experiment 2. The observers were shown displays in the side view and depth view without perspective distortions (as well as with them, in each case). The change in optical components without changes in shape tested the effect of components on recognition.

Experiment 2 was designed to test which of these possibilities was responsible for the decrement in performance in the depth-view condition. Displays were created that represented all combinations of viewing orientation and distortion, yielding four experimental conditions. In the first condition, the observers had to identify events that were presented from the side and had the normal side-view speed profiles (the side-with-side condition). In the second condition, the events were shown from the side, but with the distorted speed profiles from the original depth-view condition (the side-with-depth condition). Condition 3 consisted of displays shown in the depth view, but with the undistorted speed profiles from the original side-view condition (the depth-with-side condition). The fourth condition showed the events in depth and with the distorted speed profiles (the depth-with-depth condition). The first and fourth conditions were exactly the same events as those used in the side-view and depth-view conditions of Experiment 1. The second and third conditions used the speed profiles from the other viewing orientation. By comparing performance in the four conditions, we determined whether the optical distortions or merely the components were producing the decrement in performance in the depth-view condition of Experiment 1. If the distortions were making it more difficult to discriminate events regardless of the components, performance should be worse in Conditions 2 and 4 than in Conditions 1 and 3. If the change in compo-

nents was making it more difficult, regardless of the distortion of the speed profiles, performance should be worse in Conditions 3 and 4 than in Conditions 1 and 2.

Method

Participants. Seventeen adults participated in this experiment. The observers were recruited by e-mail and posted signs and were paid $6 per hour. The observers were told they would receive a $20 bonus if they had one of the two highest accuracies in the study, as a motivation to do well.

Design and Procedure. The same five events as those in Experiment 1 were used in this experiment. This experiment consisted of four conditions. Two of those conditions were the side-view and depth-view conditions from Experiment 1. For this experiment, these conditions were called the side-with-side (side view with the normal side-view speed profiles) and the depth-with-depth (depth view with the distorted depth-view speed profiles) conditions. Two additional conditions were a side-with-depth condition, in which the events were presented from the side, but with the distorted depth-view speed profiles, and a depth-with-side condition, in which the events were presented from the depth view, but with the nondistorted side-view speed profiles.[4]

Experiment 2 took place over 4 days. Each of the four conditions was presented on a different day. The order of the conditions was counterbalanced across observers. The structure of the session each day was exactly the same. The observers were given the opportunity to familiarize themselves with the five events from that day's condition. They were allowed to look at each display as many times as they wanted. When they had finished, they moved to the testing phase. The observers were shown one of the five events and had to identify which one it was (A, B, C, D, or E).

The observers received feedback about their performance after each judgment during the testing phase. At no time did the observers make judgments without receiving feedback. This tested the possibility that performance in Experiment 1 was worse in the depth view because the observers did not receive feedback about their performance in that condition. The session consisted of blocks of 25 trials each. Five speed profiles were shown in a random order five times each. After each block, the observers were told what their overall accuracy was for that block. The session continued as long as the observers' performance kept improving on subsequent blocks (i.e., they continued until their performance reached asymptote). The asymptote was determined by taking the mean accuracy of the previous three blocks (e.g., Blocks 3, 4, and 5) and comparing that with the mean accuracy of the previous group of three (Blocks 2, 3, and 4). This was calculated after each block. If the difference in the mean accuracy was calculated to be smaller than 3% twice in a row, it was determined that performance had leveled off, and the session was ended. If the observer's performance still had not leveled off after nine blocks, the session was terminated. This procedure was the same for each of the four conditions. Final accuracies were compared across all of the conditions.

Results and Discussion

As in Experiment 1, both percentage correct and d' statistics were calculated.[5] The mean percentages correct (and standard deviations) for the final three blocks for each condition are as follows: side with side, 93.10% (8.38); side with depth, 85.53% (8.50); depth with side, 83.14% (12.55); and depth with depth, 78.20% (14.42). Performance in every condition was well above chance (20%). As in Experiment 1, the data were transformed using a logit($\hat{p}$) to normalize the distributions for analysis. As was confirmed by a t test, the depth-with-depth condition had a significantly higher accuracy than the depth-

view condition in Experiment 1, where there was no feedback provided [$t(31) = 6.55, p < .01$]. This showed that feedback was partly responsible for the difference in results in Experiment 1. If lack of feedback was the only factor contributing to the decrement in performance in the depth-view condition in Experiment 1, there should have been no difference in performance among the four conditions in Experiment 2 (since feedback was provided in all conditions). To test for this and the remaining two hypotheses, a 2 × 2 ANOVA testing viewing orientation (side view vs. depth view) and speed profile type (undistorted vs. distorted) was conducted. There was a significant main effect for viewing orientation, showing that the observers performed significantly better in the side-view conditions than in the depth-view conditions [$F(1,15) = 5.23, p < .01$]. This supported the hypothesis that the change in optical components in the depth displays caused a decrement in performance in the depth-view condition in Experiment 1. However, the ANOVA also revealed a main effect for speed profile type [$F(1,15) = 6.04, p < .01$]. Performance was significantly better in the conditions with the undistorted speed profiles than in the conditions with the distorted profiles. This supported the hypothesis that the perspective distortions were also causing a decrement in performance in the depth-view condition in Experiment 1. The interaction was not significant [$F(1,15) = 3.93, p > .05$], indicating that the effect of perspective distortions was not dependent on the optical component (viewing orientation) or vice versa.

In addition to the difference in accuracies for each condition, we examined the potential difference in the learning rates for each condition. The average number of blocks the observers took before their performance leveled off for each condition was 6.41 (side with side), 7.12 (side with depth), 6.71 (depth with side), and 7.29 (depth with depth). The data were transformed using a square-root transformation (Johnson & Wichern, 1998). A 2 × 2 ANOVA testing viewing orientation and type of speed profile (undistorted or distorted) revealed no significant main effect of optical component [$F(1,15) = 0.27, p > .05$] or of type of speed profile [$F(1,15) = 2.26, p > .05$] and no significant interaction [$F(1,15) = 0.003, p > .05$]. Thus, there was no difference in the learning rates for each condition, just in their final accuracies. Finally, pairwise d' statistics were computed for each display in each condition. In every condition, the d's were well above 1.0, indicating that with feedback, the observers had good sensitivity and did not reliably confuse any two displays (see Tables 3 and 4 for a complete list of all d' comparisons).

GENERAL DISCUSSION

Muchisky and Bingham (2002) showed that people can use speed profiles as information for identifying events that travel along a straight path. The purpose of Experiment 1 was to examine whether or not observers exhibit constancy in event identification. We investi-

Table 3
d' Calculated in Four Viewing Conditions for Each Event Versus the Remaining Four Combined in Experiment 2

	d'			
Event Contrast	Side w/Side	Side w/Depth	Depth w/Side	Depth w/Depth
Asymmetric vs. others	3.19	2.72	2.44	2.26
Bimodal vs. others	3.31	2.46	2.51	2.09
Constant vs. others	3.42	2.36	2.30	2.08
Harmonic vs. others	3.52	2.66	2.45	1.94
Stretched vs. others	3.13	2.43	2.18	1.98
Mean sensitivity	3.31	2.53	2.38	2.07
Median sensitivity	3.31	2.46	2.44	2.08

gated whether observers could recognize events that were familiar in a side view when these events were viewed from different perspectives. The results showed that the observers could still reliably identify the speed profiles despite perspective transformations of the shapes of the events (resulting in very different optical forms in the depth-view, as compared with the side-view, condition). They could also reliably identify the speed profiles despite the change in optical components (expansion/contraction vs. translation) that occurred when the events were viewed in depth. Nevertheless, performance was significantly worse in the oblique-view condition than in the side-view condition and was worse again in the depth-view condition than in the oblique-view condition.

In Experiment 2, we examined the factors that contributed to this decrease in performance, while eliminating the role of feedback by keeping it constant across conditions. The findings of Experiment 2 showed that the change in optical components (translation vs. expansion/contraction) and the perspective transformations both contribute to the decrement in performance found in the depth-view condition in Experiment 1, although the effect of one factor did not depend on the other. This explains why we obtained intermediate performance from the observers in the oblique-view orientation. Perfor-

mance in the oblique-view condition was better than that in the depth-view condition because the image translation component was present in the oblique-view condition, but not in the depth-view condition. Average optical velocities were simply greater in the oblique-view condition due to the image translation component, and this yielded superior discrimination performance. Experiment 2 showed that the perspective transformations also yielded a drop in performance, explaining why performance declined in the oblique-view, as compared with the side-view, perspective. Thus, event recognition is comparable to object recognition in exhibiting a view specificity effect. Finally, of course, we found that feedback brought all performance levels up to about 80% correct or better and, in all cases, this level of performance was reached in about 150–175 trials. So the differences found in Experiment 1 were not only a result of differences in feedback.

Two of the perspectives that we tested—namely, from the side and in depth—were selected to represent unique end points of the spectrum of possible perspectives. In contrast, the oblique view in Experiment 1 was selected as the most representative, or generic, view. It includes both optical components, which yielded redundant information about the trajectory form because both components are subject to the same perspective transformation. Despite the slight drop in recognition performance caused by the perspective changes, the observers continued to be able to recognize the given events. So the most important conclusion of these experiments is that observers are able to recognize the same event viewed from different perspectives, using trajectory forms as information, despite difficulties posed by changes in optical components and perspective distortions.

We close by noting that we have used both the terms *transformation* and *distortion* to refer to the optical effects of perspective changes. Distortion is perhaps the more appropriate term when considering whether a perspective change will impair identification. We found that it does a bit and that there is a view specificity effect. On the other hand, we also found that the ability to recognize events is preserved despite perspective changes. In fact, our predominant result was constancy of identification, and of course, perspective transformations are required to allow these events to be recognized from various perspectives. That is, perspective changes ultimately yield informative transformations, rather than misleading distortions. The technique that we used in Experiment 2 to investigate the effect of the perspective transformations on discrimination performance actually yielded some rather different-looking events. That is, the depth-view forms applied to side-view events and, likewise, the side-view forms applied to depth-view events are real distortions that changed the event identities. The perspective transformation of the depth view or the oblique view impaired identification performance slightly by rendering the forms somewhat less discriminable, and it was this potential effect that we tested in Experiment 2. Never-

Table 4
d' Calculated in Four Viewing Conditions for Each Pair of Events in Experiment 2

	d'			
Event Contrast	Side w/Side	Side w/Depth	Depth w/Side	Depth w/Depth
Asymmetric/bimodal	3.66	2.93	2.61	2.58
Asymmetric/constant	5.16	3.74	3.84	3.88
Asymmetric/harmonic	4.91	4.78	4.26	3.46
Asymmetric/stretched	3.03	2.91	2.58	2.38
Bimodal/constant	3.56	2.01	2.76	2.32
Bimodal/harmonic	4.77	4.65	4.34	3.08
Bimodal/stretched	4.57	4.57	3.86	3.19
Constant/harmonic	3.66	3.25	2.59	2.14
Constant/stretched	4.01	3.67	3.31	3.28
Harmonic/stretched	3.80	2.36	2.38	2.12
Mean sensitivity	4.15	3.51	3.25	2.84
Median sensitivity	3.91	3.46	3.04	2.83

theless, the primary effect of perspective transformations was to provide information that allowed events to be recognized from different perspectives, allowing event recognition to exhibit constancy.

REFERENCES

BINGHAM, G. P. (1987a). Dynamical systems and event perception: A working paper. Pt. I. *Perception–Action Workshop Review*, **2**, 4-14.

BINGHAM, G. P. (1987b). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 155-177.

BINGHAM, G. P. (1995). Dynamics and the problem of visual event recognition. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 403-448). Cambridge, MA: MIT Press.

BINGHAM, G. P., ROSENBLUM, L. D., & SCHMIDT, R. C. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1473-1493.

CUTTING, J. E., & KOZLOWSKI, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, **9**, 353-356.

FRYKHOLM, G. (1983). *Action, intention, gender, and identity, perceived from body movement* (Acta Universitatis Upsaliensis: Abstracts of Uppsala Dissertations from the Faculty of Social Sciences, No. 31).

GILDEN, D. L., & PROFFITT, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 372-383.

GILDEN, D. L., & PROFFITT, D. R. (1994). Heuristic judgment of mass ratio in two-body collisions. *Perception & Psychophysics*, **56**, 708-720.

JOHANSSON, G. (1950). *Configurations in event perception.* Uppsala, Sweden: Almqvist & Wiksell.

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

JOHANSSON, G. (1976). Spatio-temporal differentiation and integration in visual motion perception: An experimental and theoretical analysis of calculus-like functions in visual data processing. *Psychological Research*, **38**, 379-393.

JOHNSON, R. A., & WICHERN, D. W. (1998). *Applied multivariate statistical analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

KAY, B. A., KELSO, J. A. S., SALTZMAN, E. L., & SCHÖNER, G. (1987). Space–time behavior of single and bimanual rhythmical movements: Data and limit cycle model. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 178-192.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

MUCHISKY, M. M., & BINGHAM, G. P. (2002). Trajectory forms as a source of information about events. *Perception & Psychophysics*, **64**, 15-31.

NEWELL, F. N., ERNST, M. O., TJAN, B. S., & BÜLTHOFF, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, **12**, 37-42.

ORBAN, G. A., DE WOLF, J., & MAES, H. (1984). Factors influencing velocity coding in the human visual system. *Vision Research*, **24**, 33-39.

PEROTTI, V. J., TODD, J. T., LAPPIN, J. S., & PHILLIPS, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics*, **60**, 377-388.

RUNESON, S. (1977). *On visual perception of dynamic events.* Uppsala, Sweden: University of Uppsala.

RUNESON, S. (1983). *On visual perception of dynamic events* (Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensia, No. 9). Stockholm: Almqvist & Wiksell. (Original work published 1977).

RUNESON, S. (1995). Support for the cue-heuristic model is based on suboptimal observer performance: Response to Gilden and Proffitt (1994). *Perception & Psychophysics*, **57**, 1262-1273.

RUNESON, S., & FRYKHOLM, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, **112**, 585-615.

RUNESON, S., & VEDELER, D. (1993). The indispensability of precollision kinematics in the visual perception of relative mass. *Perception & Psychophysics*, **53**, 617-632.

SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

TARR, M. J., PEPPER, W., HAYWARD, W. G., & GAUTHIER, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, **1**, 275-277.

WICKELGREN, E. A., & BINGHAM, G. P. (2001). Infant sensitivity to trajectory forms. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 942-952.

NOTES

1. See Bingham (1987a and 1987b) for initial attempts to attack this problem.

2. Nevertheless, the time dimension may become relevant in some cases in which events are distinguished by periods during which motion ceases (e.g., a hopping frog vs. a bouncing ball). There is no information in a phase portrait about how long the trajectory remains at zero velocity once zero velocity has been achieved.

3. The response matrices combining all 16 observers in the side-view, oblique-view, and depth-view conditions in Experiment 1 are shown in Tables A1, A2, and A3, respectively, in Appendix A.

4. Note that the depth and the oblique speed profiles had the same shape (see Figures 3 and 4), so although we chose to use the depth speed profiles in this experiment, the oblique profiles could have been substituted and still yielded the same results, since the kinematic forms were the same.

5. The overall response matrices for the side-with-side, side-with-depth, depth-with-side, and depth-with-depth conditions in Experiment 2 are shown in Tables B1, B2, B3, and B4, respectively, in Appendix B.

APPENDIX A

Table A1
Confusion Matrices for the Side-View Condition in Experiment 1

	Observers' Responses				
Events Presented	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	83	7	0	0	6
Bimodal	1	91	4	0	0
Constant	1	2	92	1	0
Harmonic	1	0	8	87	0
Stretched	11	3	0	9	73
Total responses	97	103	104	97	79

Table A2
Confusion Matrices for the Oblique-View Condition in Experiment 1

	Observers' Responses				
Events Presented	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	59	9	4	2	22
Bimodal	24	44	11	4	13
Constant	11	11	57	7	10
Harmonic	10	7	9	42	28
Stretched	13	6	3	19	55
Total responses	117	77	84	74	128

Table A3
Confusion Matrices for the Depth-View Condition in Experiment 1

	Observers' Responses				
Events Presented	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	45	16	11	10	14
Bimodal	13	28	27	21	7
Constant	10	14	53	12	7
Harmonic	5	5	10	58	18
Stretched	15	7	5	36	33
Total responses	88	70	106	137	79

APPENDIX B

Table B1
Confusion Matrices for the Side-With-Side Condition in Experiment 2

	Observers' Responses				
Events Presented	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	484	22	3	3	33
Bimodal	13	488	36	3	5
Constant	2	10	513	14	6
Harmonic	4	6	22	500	13
Stretched	34	6	8	16	481
Total responses	537	532	582	536	538

Table B2
Confusion Matrices for the Side-With-Depth Condition in Experiment 2

	Observers' Responses				
Events Presented	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	503	34	14	3	51
Bimodal	43	485	69	3	5
Constant	17	80	467	29	12
Harmonic	6	8	24	497	70
Stretched	29	6	22	63	485
Total responses	598	613	596	595	623

APPENDIX B (Continued)

Table B3
Confusion Matrices for the Depth-With-Side Condition in Experiment 2

Events Presented	Observers' Responses				
	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	439	55	21	6	49
Bimodal	41	456	56	10	7
Constant	7	29	436	77	21
Harmonic	10	5	31	483	41
Stretched	43	19	22	82	404
Total responses	540	564	566	658	522

Table B4
Confusion Matrices for the Depth-With-Depth Condition in Experiment 2

Events Presented	Observers' Responses				
	Asymmetric	Bimodal	Constant	Harmonic	Stretched
Asymmetric	457	58	20	19	66
Bimodal	43	457	72	29	19
Constant	15	57	455	71	22
Harmonic	20	30	65	435	70
Stretched	52	35	25	76	432
Total responses	587	637	637	630	609