Perception & Psychophysics
2002, 64 (2), 244-265

The duration of 3-D form analysis in transformational apparent motion

PETER ULRIC TSE and NIKOS K. LOGOTHETIS
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Transformational apparent motion (TAM) occurs when a figure changes discretely from one configuration to another overlapping configuration. Rather than an abrupt shape change, the initial shape is perceived to transform smoothly into the final shape as if animated by a series of intermediate shapes. We find that TAM follows an analysis of form that takes 80–140 msec. Form analysis can function both at and away from equiluminance and can occur over contours defined by uniform regions as well as outlines. Moreover, the forms analyzed can be 3-D, resulting in motion paths that appear to smoothly project out from or into the stimulus plane. The perceived transformation is generally the one that involves the least change in the shape or location of the initial figure in a 3-D sense. We conclude that perception of TAM follows an analysis of 3-D form that takes ~100 msec. This stage of form analysis may be common to both TAM and second-order motion.

When two nonoverlapping figures are flickered in succession within a certain range of spatiotemporal offsets (Korte, 1915), they appear to comprise a single object jumping rigidly back and forth in *translational apparent motion*. Because no object actually moves in the world, the appearance of motion must be created by the visual system. Exner was the first to scientifically investigate the properties of translational apparent motion (Exner, 1875, 1888; also called *beta motion*; see Steinman, Pizlo, & Pizlo, 2000; Wertheimer, 1912/1961, 1923/1938). He showed that observers can perceive an illusory jump between two sparks of light even when the positions of those sparks cannot be visually discriminated. This proved that motion perception is a primary sensation, not reducible to and not dependent on the prior computation of spatial displacement or temporal intervals. Whereas Exner tried to explain apparent motion in terms of the workings of the eye, Wertheimer (1912/1961) argued that apparent motion demonstrates the cortically constructed nature of perception. He and other Gestalt psychologists (e.g., Koffka, 1922, 1935; Köhler, 1959) described numerous grouping procedures that could not be explained by reductionistic approaches to mental phenomena (e.g., the view that perception is built up from units analogous to atoms; see, e.g., Russell, 1918/1965; Wittgenstein, 1921/1961; or, later, the view that there is no mental processing at all; see, e.g., Skinner, 1935, 1948; J. B. Watson, 1913). As a consequence of Gestalt insights, most modern psychologists now assume that consciously experienced vision is the product of extensive unconscious information processing, including grouping procedures. In particular, the perception of apparent motion is thought to be the product of neuronal circuitry that interprets successive discrete inputs as having arisen from one or more continuously moving objects in the world.

The most important conceptual issue in the field of translational apparent motion has been called the *correspondence problem*. When there are several nonoverlapping figures in each of several successive images, the visual system is faced with an ambiguity. It must determine which figure in the first image corresponds to which in the second before motion between corresponding figures can be constructed. There are usually multiple possible matches that can be made between figures in successive images. Many researchers have tried to determine what stimulus factors drive the visual system to make the matches that it does. Studies using translational apparent motion sequences as a probe have shown that form analysis plays little role in solving the correspondence problem (e.g., Baro & Levinson, 1988; Burt & Sperling, 1981; Cavanagh, Arguin, & von Grünau, 1989; Dawson, 1991; Kolers, 1972; Kolers & Pomerantz, 1971; Kolers & von Grünau, 1976; Krumhansl, 1984; Navon, 1976, 1983; Victor & Conte, 1990). Indeed, low spatial frequencies appear to count for more than shape (Green, 1986b; Ramachandran, Ginsburg, & Anstis, 1983), and the most potent factor is found to be spatiotemporal proximity. Although there are examples in which shape or color do play a role in matching (e.g., Green, 1986a, 1989; Shechter, Hochstein, & Hillman, 1988), the effect of these factors is revealed only when the much stronger factor of proximity is carefully controlled for. This tendency to match figures across successive displays on the basis of proximity irrespective of shape or color correspondence has been called the *nearest neighbor principle* (Ullman, 1978, 1979; Wertheimer, 1912/1961). Nearest neighbor matching is expressed by

This work was supported by McDonnell-Pew Grant 98-49CNS and by the Max Planck Society. The authors thank Ken Nakayama and Patrick Cavanagh for helpful discussions. Correspondence should be addressed to P. U. Tse, Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hanover, NH 03755.

Copyright 2002 Psychonomic Society, Inc.

most models of visual motion processing (Adelson & Bergen, 1985; Barlow & Levick, 1965; Hassenstein & Reichardt, 1956; Reichardt, 1961; van Santen & Sperling, 1984) that base matches on the nature of luminance contrast changes (i.e., "motion energy") over brief intervals.

The seeming indifference of motion processing to form information in apparent motion meshed well with neurophysiological findings that the low spatial, high temporal resolution magnocellular system, specialized for the processing of motion and spatial relationships, responds poorly to contours and boundaries defined only by color contrast (e.g., Livingstone & Hubel, 1987; Ramachandran & Gregory, 1978; but see also Albright, 1992; Dobkins & Albright, 1994; Saito, Tanaka, Isono, Yasuda, & Mikami, 1989). The high spatial, low temporal resolution parvocellular-interblob system, specialized for form analysis, meanwhile, was found to respond poorly to motion (e.g., Livingstone & Hubel, 1987; Schiller, 1991). Livingstone and Hubel suggested that the magnocellular system may specialize in the rapid analysis of depth and brightness, since these can be performed at low spatial resolution, unlike form or color analysis, which require high resolution. The segregation of motion and form processing was further supported by evidence that spatial processing and object recognition are handled in separate cortical pathways (Ungerleider & Mishkin, 1982). Motion and form are processed in at least partially distinct anatomical subregions of these pathways, with the middle temporal region (MT in macaques, also called V5 in humans) playing a central role in motion processing (e.g., Albright, 1984; Newsome & Pare, 1988; Salzman, Murasugi, Britten, & Newsome, 1992), and occipitotemporal and inferotemporal regions playing a central role in form analysis (e.g., Pasupathy & Connor, 1999) and object recognition (e.g., Logothetis & Sheinberg, 1996). Although form and motion processing may proceed independently in the ventral and dorsal pathways, there is evidence that these streams influence each other. The form and motion pathways appear to converge in area STP (Baizer, Ungerleider, & Desimone, 1991), and cells have been discovered that appear to be tuned to motions of highly specific forms, such as the motions of the human body (Oram & Perrett, 1994; Perrett, Harries, Mistlin, & Chitty, 1990).

In 1991, a new type of motion illusion was discovered that seemed difficult to explain using standard models of motion processing (Hikosaka, Miyauchi, & Shimojo, 1991, 1993a, 1993b; for partial precedents, however, see Bundesen, Larsen, & Farrell, 1983, Farrell & Shepard, 1981, Hartmann, 1923, Kanizsa, 1951, 1979, Kenkel, 1913, Kolers & Pomerantz, 1971, Orlansky, 1940, and Shepard, 1984). Several investigators examined simple two image motion sequences in which one shape was replaced all at once by another one that overlapped the first. In the simplest and first version of this phenomenon (Hikosaka et al., 1991), a spot flashed, disappeared, and then was replaced all at once by a line whose end overlapped or abutted the position previously occupied by the spot. Even though there were only two figures in this motion display—namely, one spot replaced by a single line—the line appeared to shoot out of the spot as if it were animated by a sequence of lines of intermediate length. Hikosaka et al. suggested that this phenomenon is due to attention. In particular, they argued that *illusory line motion*, as they termed it, could be explained by the principle of attentional *prior entry* (Stelmach & Herdman, 1991; Stelmach, Herdman, & McNeil, 1994; Sternberg & Knoll, 1973; Titchener, 1908), which states that visual information near an attended locus is processed more quickly than information elsewhere. According to their model, when the initial spot appears, an attentional gradient forms around the spot; when the line appears all at once, portions of the line nearer the attended spot enter a motion detecting mechanism before portions of the line that lay farther away from the attended location. Other authors (Downing & Treisman, 1995, 1997; Kawahara, Yokosawa, Nishida, & Sato, 1996; Tse & Cavanagh, 1995; Tse, Cavanagh, & Nakayama, 1996, 1998), using more complex first and second image shapes, challenged the attentional prior entry account. They argued instead that illusory line motion was actually a variant of translational apparent motion (e.g., Anstis, 1980; Braddick, 1980; Dawson, 1991; Wertheimer, 1912/1961). These authors found that attentional factors could not entirely account for the line motion effect, because stimulus-driven factors, such as contour relationships, can influence the perceived direction of motion.

Tse and colleagues (Tse & Cavanagh, 1995, 2000; Tse et al., 1996, 1998) coined the term *transformational apparent motion* (TAM) to contrast smooth apparent shape changes with translational apparent motion. Both classes of apparent motion result from constructive processes that infer how an object must have changed in the world given discrete input. The "line motion" of Hikosaka et al. (1991) is then just a special case of apparently smooth surface and object shape changes. In TAM sequences, one shape is replaced all at once by a different overlapping shape. The first shape appears to undergo a smooth and continuous transformation into the second shape as if animated by a series of intermediate shapes. This illusion of a smooth shape change is, phenomenally speaking, the defining characteristic of TAM that sets it apart from translational apparent motion. TAM is depicted in Figure 1 (an example of TAM can be on the first author's homepage at the Department of Psychology and Brain Sciences, Dartmouth College).

TAM demonstrated that indifference to form is an artifact of using translational apparent motion as a probe, because, in translational apparent motion, successive figures do not overlap (Tse et al., 1998). When successive figures do overlap, the visual system is confronted with a deeper problem than the problem of matching a figure in one scene to itself in the next. It must specify what counts as a figure in each scene before matches can be made. When figures overlap within or between scenes, as in Figure 1, a stage of form analysis must decide what counts as a figure before it can be matched to its (perhaps trans-

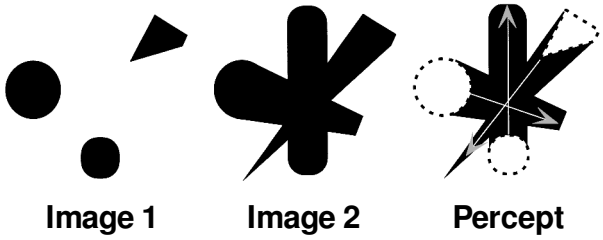

Figure 1. A typical example of transformational apparent motion. When Image 2 replaces Image 1, the Image 1 figures undergo smooth motions into their final Image 2 shapes, as shown using dotted figures and arrows in Image 3. When Image 2 appears, the visual system is faced with a form analysis problem that must be solved before figures can undergo apparent motion into the corresponding portions of the single shape in Image 2.

formed) self in the next image. This stage of form analysis must be rapid in order for motions to be perceived veridically, without excessive delay between perceived motions and the motions occurring in the world. Although this "parsing problem" logically precedes the correspondence problem (Tse et al., 1998), its solution need not temporally precede the solution to the correspondence problem, because neuronal circuits could be designed that solve both problems simultaneously. In particular, 2-D or 3-D figures may best be conceived as spatiotemporal, rather than just spatial, entities that are parsed, tracked, and matched over time. In this light, the parsing or form analysis problem and the correspondence problem are continually being addressed for each new image with reference to recent past images. This point of spatiotemporal form and motion processing will be discussed in more detail in the General Discussion section. Note that the parsing problem does not even arise in translational apparent motion displays because there is no ambiguity regarding what counts as a figure within any single image of a translational apparent motion sequence.

The aim of the present research was to determine the basic spatiotemporal characteristics of the stage of form analysis involved in generating TAM. At one extreme, the stage of form analysis could be the same as that available with unlimited viewing of static images, involving edge detection, contour formation, grouping, segmentation, 3-D analysis, and object recognition. At the other extreme, the stage of form analysis might involve no more than the shape of receptive fields of neurons contributing to motion energy detecting networks. It is unlikely that either extreme is used to generate TAM. On the one hand, we do not need to recognize an object to be able to see it move. On the other hand, empirically observed low-level motion-tuned receptive fields are not tuned to either local or global form but instead are sensitive to the spatial frequency content of the image and are generally only successful at discriminating motion very locally, over a fraction of the spatial wavelength to which they are tuned (Cheng, Hasegawa, Saleem, & Tanaka, 1994; Zeki, 1974). Many researchers have therefore assumed that form analysis involves grouping and parsing procedures that start with edge detection and build more complex contour and shape representations from these primitives (e.g., Beck, 1982; Julesz, 1984a, 1984b; Marr, 1982). Because form analysis in the motion pathway is likely to involve intermediate representations at the contour and surface level, the present research focused on the role of contour and surface relationships among successive figures in TAM.

Although motion energy and attention (e.g., Hikosaka et al., 1993a, 1993b) can influence the matching process in TAM, and attention plays a role in the matches that underlie translational apparent motion (Dick, Ullman, & Sagi, 1987; Horowitz & Treisman, 1994; Ivry & Cohen, 1990; Stelmach et al., 1994), these two factors will not be considered here. The properties of form analysis in the motion processing system can be studied only when the potentially confounding factors of motion energy and attention are controlled for. We are interested in the form-driven processing that underlies TAM. This processing cannot be explained by motion energy models of TAM (e.g., Zanker, 1997). A typical example of TAM that violates all motion energy models that link the centroid of one luminance blob to the nearest neighbor centroid in the next image is shown in Figure 2. Here, motion is perceived to commence away from the smaller cue, although the centroid of the new region in Image 2 is much closer to the cue on the right. This implies that the motion percept results from image information other than motion en-

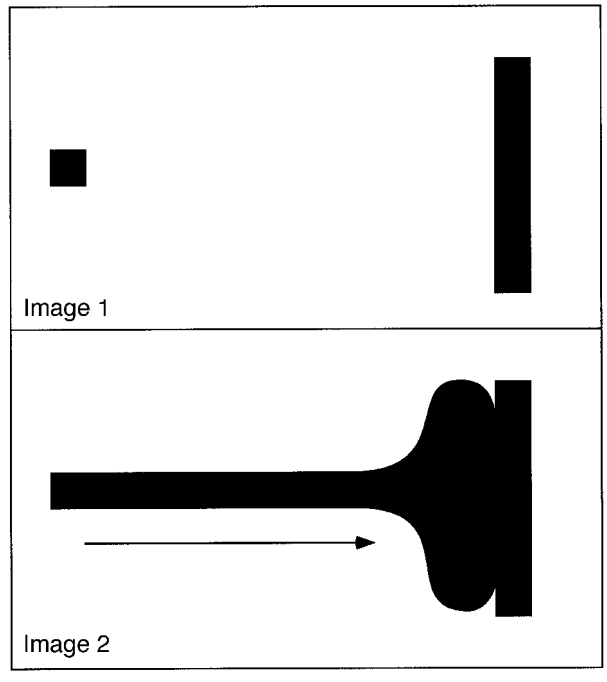

Figure 2. Transformational apparent motion can proceed in a direction that contradicts all models of apparent motion based solely on motion energy.

ergy. In particular, the direction of TAM here is driven by contour relationships between and within Images 1 and 2.

We use TAM because TAM is better than translational apparent motion for addressing questions of form analysis in the motion pathway. Using translational apparent motion as a probe to address questions of form analysis requires that biases introduced by the low-level motion processing system be eliminated before effects of form-based matching can be seen. That is, form-based matching occurs only in translational apparent motion when matching on the basis of the nearest neighbor principle has been controlled for (Green 1986a, 1989; Shechter et al., 1988) or when the the low-level motion system's response is attentuated by an interstimulus interval (ISI) of sufficient length (Shiffrar & Freyd, 1993; Ternus, 1926/1938). In TAM, however, matching does not primarily take place on the basis of the nearest neighbor principle (Tse et al., 1998). Rather, it takes place on the basis of grouping procedures that link successive figures as transformations in the shape of a single figure over time. TAM is an ideal tool for addressing questions of form analysis in the motion pathway, because illusory transformational motion will only be perceived when a new shape has been analyzed and matched to a shape from the previous image, assuming motion energy and attentional contributions to TAM are controlled for.

We ask two primary questions about form–motion interactions, using TAM as a probe. First, how long must form information be present in order to bias motion perception in TAM stimuli? The answer to this question should place an upper bound on how long form analysis takes, because only after a form has been analyzed can it be matched to a form existing in the previous image, such that a transformation between those two forms is perceived. We use simple 2-D and 3-D motion sequences to establish that form information must be available for 80–140 msec in order to bias TAM. The second question we ask concerns the type of form information processed by the form processor whose output biases the perceived direction of TAM. We find that correspondence in TAM sequences can be determined by the sharing of continuous contours among successive, overlapping stimuli. This is true whether the contours are defined by solid shapes or outlines at or away from equiluminance. Finally, we find that the 3-D form perceived in one image of a multiimage TAM sequence can influence the perceived direction of motion in a later image of that sequence. This suggests that it is perceived 3-D form that biases the motion path generated by the motion processing stream and that high-level motion processing itself determines motion paths in a 3-D coordinate frame.

A brief description of each experiment and our main results are summarized in the Appendix. Experiments 3 and 5 were the most important, because they established that 3-D form analysis takes ~100 msec. The other experiments were control experiments.

EXPERIMENT 1

In order to address the central question of the speed of form analysis in the motion pathway and to control for confounding factors due to motion energy and attention, a "limiting case" of the well-known quartets stimuli was used. In a traditional quartets stimulus (Ramachandran et al., 1983; Wertheimer, 1912/1961), two nonoverlapping figures, such as disks, occupy the diagonally opposite corners of an imaginary square in Image 1 and occupy the other two corners in Image 2 of a two-image translational apparent motion sequence. Because proximity is controlled for when observers fixate a fixation point at the center of this imaginary square, observers tend to see horizontal and vertical motions with equal probability when there is a brief blank ISI between the images. The stimulus used here was the limiting case of a quartets stimulus, because diagonally opposite quadrants of an imaginary square were filled, forming smaller equiluminant red or green squares without overlap in Images 1 and 2, as shown in Figure 3. The squares undergoing apparent motion could not get any closer to one another without overlapping across images.

Experiment 1 was a control experiment designed to test whether matching could take place on the basis of color alone. If matching takes place on the basis of color rela-

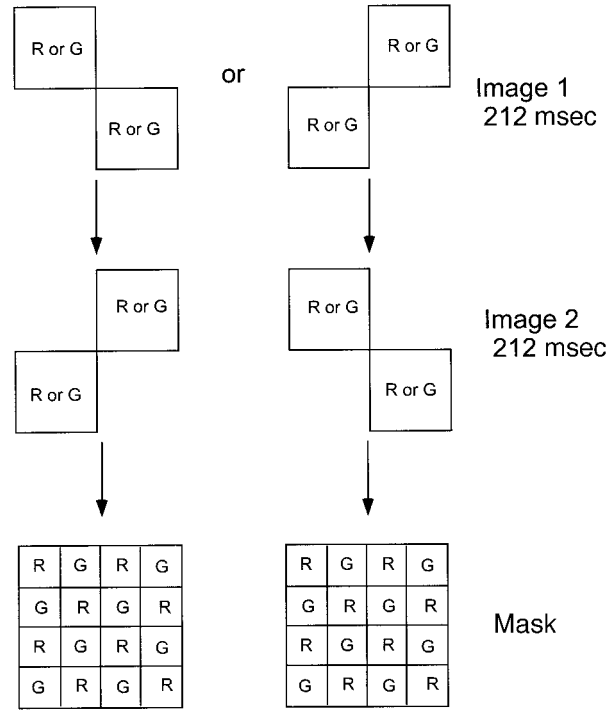

Figure 3. The basic experimental paradigm for Experiment 1 was a "limiting case" of the quartets stimulus, in that the elements could not approach one another any more closely without overlapping. Two diagonally opposite quadrants of an imaginary square were occupied by smaller squares. The other two quadrants were occupied in Image 2, followed by a mask.

tionships among figures in successive images, then color cannot be used to define forms in order to study form-based matching in TAM. This is because matching might then take place on the basis of color and not form. If matching does not take place on the basis of color, we can use color as a means of defining form in order to determine how form biases perceived motion paths in TAM.

Method

Every possible combination of quadrant colors was tested 10 times in Experiment 1, for a total of 320 trials per observer (320 trials = 10 × 4 [possible color combinations for Image 1: red–red, red–green, green–red, green–green] × 4 [possible color combinations for Image 2] × 2 [Image 1 positions]). Presentation order was randomized. Images 1 and 2 had a 0 ISI (see Figure 3). Of the 320 trials, the 80 trials of interest were those in which Images 1 and 2 each had one red and one green quadrant, because, in these trials, matching could have taken place on the basis of a unique color correspondence between images. This set of 80 was not shown in isolation in order to control for the possibility of attentionally tracking on the basis of color. On 240 of 320 trials, any strategy based on color tracking or cognitive inference would not have been useful, because, in these cases, there was not a single red and green square in Images 1 and 2.

Observers. The observers were obtained from the Max Planck Institute observer pool and were paid. All had normal or corrected-to-normal vision. All observers were naive regarding the purpose of the experiments. The observer sat on a chair in a small cubicle entirely enclosed by dark curtains and rested the chin in a chinrest. The distance from the observer's eyes to the screen was 57 cm.

Stimulator. The visual stimulator was a dual-processor Pentium II workstation running Windows NT (Intergraph Corp., Huntsville, AL) equipped with a VX25 graphics subsystem. The screen resolution was 1,152 × 864 pixels, and the image rate was 85 Hz. All image generation was in 24-bit truecolor, using hardware double buffering. The stimulation software was written in C and was based around Microsoft's OpenGL 1.1 implementation (with the client driver specific for Intergraph hardware). The Tool Command Language (Ousterhout, 1994) was embedded within the stimulation software and was used for scripting. External control of the visual stimulator was achieved by using a high-speed ethernet connection between the real-time control PC and the dedicated visual workstation. Exact timing was verified by generating framebuffer swap synchronization pulses with a digital timer I/O card installed in the stimulator (ACL8454, Circuit Specialists, Inc.).

Stimuli. Occupied quadrants in Images 1 and 2 (see Figure 3) subtended 4.4° × 4.4°. Image 1 and Image 2 were presented for 212 msec each. The background was black. The red and green values were set to objective equiluminance using a photometer. The red (CIE, x = .64, y = .33) was set to its maximum value, and the green value (CIE, x = .29, y = .60) was lowered until it had the same luminance of 11.90 cd/m^2. The background had a luminance value of less than 0.01 cd/m^2.

Mask. The mask had two parts. First, the whole screen turned white for 212 msec. Second, a 8.8° × 8.8° checkerboard pattern made up of 16 alternating red and green squares appeared for 212 msec. The red and green values were those used in Images 1 and 2.

Fixation point. The fixation point was a 0.1° × 0.1° yellow square. Eye movements were not monitored during the task but were monitored by the experimenter in practice trials. All observers were able to maintain fixation and reported no trouble doing so. The fixation point was always present on the screen between trials. The beginning of a trial was indicated by a 106-msec offset of the fixation point. The fixation point then reappeared for 106 msec, followed by Image 1 of the quartets sequence. When the quartets stimuli were present, the fixation point was instantaneously replaced by the vertices of the visible quadrants of the imaginary square defining the quartets stimulus. Immediately after offset of the last image of the mask, the fixation point reappeared in its previous position. After the observer responded, the intertrial interval, in which only the fixation point was present, lasted for 1,000 msec. If the observer did not respond, the next trial did not appear.

Procedure. Eight observers responded in a two-alternative forced-choice paradigm. One buttonpress indicated that they had seen vertical motion, and the other indicated that they had seen horizontal motion. The observers were instructed to passively report the motion that they had seen and were told to attend to the whole stimulus configuration, rather than any single quadrant. They were led to believe that there was a real motion in the stimuli in all experiments, and their task was to detect the direction (horizontal or vertical) of this motion.

Results

The average percentage of apparent motion that was perceived in the direction that would be predicted on the basis of color matching for the subset of 80 trials in which matching on the basis of color was possible was at chance (n = 8; 51.41% ± 0.87 standard error).

Discussion

In the subset of trials in which there was a single red and a single green square in Images 1 and 2, matching on the basis of color would predict red matching to red and green matching to green. This should create an unambiguous perceived direction of motion. The observers were at chance, however, indicating that matching did not take place on the basis of color. Indeed, most observers tended to see one or the other motion again and again for many trials, indicating that there was motion hysteresis that was not broken by color matching. Once the direction of motion flipped to the opposite motion, the observers tended to see motion in this new direction for many trials. Thus, matching does not take place on the basis of color under these stimulus conditions. Experiments (Green, 1986a, 1986b, 1989) in which color has been reported to play a role in correspondence matching may have resulted from attentional tracking or cognitive inference on the basis of color, rather than from automatic correspondence matching on the basis of color per se. Here, we controlled for attentional tracking or cognitive inference on the basis of color by embedding the stimuli of interest among stimuli where such strategies would not have been effective.

EXPERIMENT 2

Experiment 2 was also a control experiment. Image 2 of a three-image apparent motion sequence was composed of two "biasing" rectangles, each of which occupied two quadrants of the imaginary square (see Figure 4). This experiment was carried out to determine the maximal Image 2 duration that allowed the impression of smooth and continual motion from Image 1 to Image 3. When Image 2 was too long, the motion from Image 1 appeared to stop at Image 2 before continuing on from Image 2 to Image 3. This "border" between a smooth and halting mo-

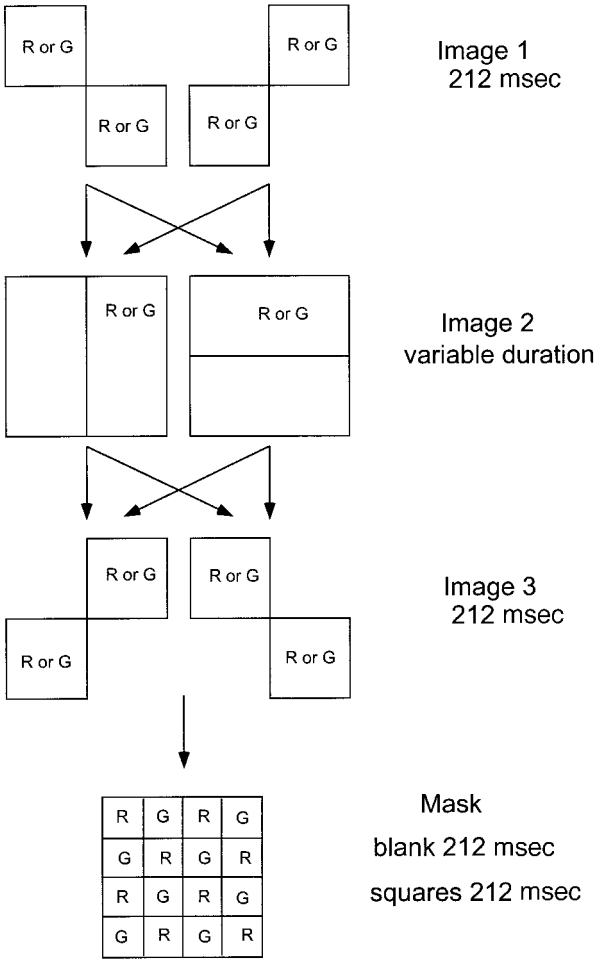

Figure 4. The two diagonally opposite quadrants occupied in Image 1 were replaced in Image 2 by two rectangles that were aligned either vertically or horizontally within the same imaginary square. The duration of Image 2 was varied. In Image 3, the opposite quadrants from Image 1 were occupied, followed by a mask.

tion between Images 1 and 3 was determined, because the observers in Experiment 3 were led to believe that there was a real, continual motion in the stimulus, whose direction they had to detect. All Image 2 durations tested in these other experiments were within the smooth motion domain, as defined by this experiment.

Method

Two naive observers who did not take part in the other experiments used the method of adjustment to determine the maximum duration of Image 2 that still allowed the motion from Image 1 via Image 2 to Image 3 to appear like a smooth, continual transformational motion. These observers were not led to believe that there was real motion in the stimulus. They were told that the motion was an illusion, consisting of three individual static images. The first image always had one red and one green square. The green (red) rectangle of the second image always overlapped the green (red) square of the first. The third image also consisted of one red and one green square. Again, the green (red) square always overlapped the green (red) rectangle of the previous image. The observers could lengthen the duration of the middle biasing image by pressing the right arrow, and they could shorten it by pressing the left arrow key. They could show one Image 1–2–3–mask sequence by pressing the up arrow key. When they had pressed the up arrow many times and felt confident that they had found the longest duration where the motion still looked smooth, they indicated this to the experimenter, and the duration was recorded. Each observer did this eight times, one for each Image 1, 2, and 3 red/green combination, and the average of these eight durations was calculated. The initial duration of Image 2 was 12 msec on four trials and 400 msec on the other four trials. Other experimental parameters were the same as those in Experiment 1.

Results

The average maximal Image 2 duration that still created the impression of a smooth motion from Image 1 to Image 3 was 251.6 ± 6.3 msec for one observer and 216.1 ± 8.7 msec for the other (average = 233.9 ± 17.7 msec).

Discussion

When Image 2 is as long as 216 msec in duration, Image 1 still appears to undergo a nonhalting transformational motion into Image 3. In Experiment 3, we varied the duration of Image 2 in order to determine the duration that can bias the perceived direction of motion from Image 1 into Image 3. If the durations tested in the following experiments (Experiments 3A, 3B, 3C, and 3D) were above 216 msec, then it could be argued that observers were not using motion as a cue in order to respond which direction of motion they had perceived but rather were using some other cue, such as the perceived orientation of the stationary rectangles in Image 2. Therefore, the Image 2 durations tested in the following four experiments were 59, 82, 106, 129, 153, and 212 msec, all of which created the impression of a smooth motion between squares in Image 1 and Image 3 (see Figure 4), such that the biasing rectangles in Image 2 were invisible as stationary rectangles.

EXPERIMENT 3

Experiments 3A, 3B, 3C, and 3D were variations on a theme and will be discussed together. If apparent motion is biased in the horizontal (vertical) direction when the Image 2 rectangles are stacked horizontally (vertically) regardless of the colors used to define those rectangles, then correspondence matching must take place on the basis of form relationships among successive images and not on the basis of color relationships.

In Experiment 3A, the first Images 1 and 3 (see Figure 4) always consisted of one red square and one green equiluminant square on a black background. Note that this is different from Experiment 1, in which Image 1 could consist of two same-colored squares. Similarly, Image 2 consisted of one red rectangle and one green rectangle. There were 32 permutations and 10 tests of each permutation for a total of 320 trials per block (320 trials = 10 × 2 [possible color combinations for Image 1] × 2 [possible color combinations for Image 3] × 2 [Image 1 positions] × 2 [Image 2 rectangle combinations; red–green, green–red] × 2 [Image 2 rectangles aligned either horizontally or

vertically]). Presentation order was randomized. Images 1 and 3 each lasted 212 msec. Image 2 had a variable duration. Each Image 2 duration (59, 82, 106, 129, 153, and 212 msec) was tested in a separate block of 320 trials. Blocks were tested in random order.

In Experiment 3B, stimuli were identical to those of Experiments 2 and 3A, except that all stimuli were equiluminant with the intermediate gray background. Real motion and translational apparent motion (e.g., Cavanagh, Boeglin, & Favreau, 1985; Ramachandran, Armel, Foster, & Stoddard, 1998) are weak at equiluminance. This experiment addressed whether TAM is perceived at equiluminance and whether form analysis can bias motion processing in this domain.

In Experiment 3C, stimuli were identical to those of Experiments 3A and 3B, except that all stimuli were black outline squares or rectangles on a white background. Experiment 3D tested white outline squares on a black background. These two experiments addressed whether TAM is perceived over outlines and whether form analysis can bias motion processing in this domain.

Method

Observers. Eight (Experiments 3A and 3B) or 7 (Experiments 3C and 3D) observers responded in a two-alternative forced-choice paradigm. They were led to believe that there was a real motion in the stimuli in all experiments, and their task was to detect the direction (horizontal or vertical) of this motion.

Stimuli. Stimuli were similar to those used in Experiment 2. Image 1 and Image 3 were presented for 212 msec each, separated by Image 2, which had a variable duration of 59, 82, 106, 129, 153, or 212 msec. These ideal times were rounded from the nearest multiple of the monitor's refresh rate of 11.76 msec (85 Hz).

In Experiment 3B, both red and green subjective equiluminance points were determined for each observer using the minimal flicker method (Anstis & Cavanagh, 1983) at approximately 43 Hz against the gray background used in the experiment. The green was a pure green, but to lower the saturation of the red square, blue and green were added in the following proportions: red = 1; green = .31; blue = .24. This was done because a pure red was much more salient than a pure green on the gray background, although both were equiluminant with that background. The luminance value for the gray background was 5.70 cd/m^2 (CIE, x = .27, y = .28) and was close to 6.43 cd/m^2 for the subjectively equiluminant red (CIE, x = .56, y = .39) and green (CIE, x = .29, y = .60) for all observers.

Mask. The red and green values of the mask in Experiments 3A and 3B were those used for the quartets stimuli in Experiments 1, 2, and 3A. The masks for Experiments 3C and 3D were also composed of 16 smaller squares arranged into a larger 8.8° × 8.8° square. In Experiment 3C, the squares were defined by black lines 1 pixel thick on a white background; in Experiment 3D, they were defined by white lines 1 pixel thick on a black background.

Fixation point. The fixation point was a 0.1° × 0.1° yellow square for all experiments, except Experiment 3C, in which it was dark blue.

Experiments 3A–3D consisted of six blocks each, one for each duration tested. Each block consisted of 320 trials, counterbalanced and randomized for every factor. Blocks were presented in random order.

Data analysis. The percentage of responses in the form-biased direction was calculated for each Image 2 duration. A Weibull curve was fit to the data in order to determine the Image 2 duration that defined the threshold where perceived motion went in the direction predicted by form biasing 75% of the time. This 75% threshold was taken as an estimate of the duration that the stage of form analysis needs in order to attain a form solution that can then bias correspondence matching.

Results

Experiment 3A. The Image 2 duration that defined the 75% threshold was 126.4 ± 3.8 msec (n = 8). This was the duration for which a stimulus had to be present in order to bias the direction of perceived motion 75% of the time. When the duration of Image 2 was 212 msec, all observers saw smooth transformational motion in the form-biased direction nearly 100% of the time (98.63% ± 0.44; n = 8). When Image 2 was only 59 msec in duration, however, performance was near chance for all observers (52.03% ± 0.88).

Experiment 3B. One observer displayed no form-based bias at equiluminance and was at or near chance for all Image 2 durations. In debriefing, this observer reported that she was unable to see the green stimulus very well, which probably accounts for her poor performance. All other observers (n = 7) saw motion in the form-biased direction nearly 100% of the time (98.71% ± 0.54) when the duration of Image 2 was 212 msec. When Image 2 was only 59 msec in duration, however, performance was near chance for these 7 observers (52.19% ± 0.38). The Image 2 duration that defined the 75% threshold for these 7 observers was 106.4 ± 7.0 msec.

Experiment 3C. The Image 2 duration that defined the 75% threshold for the 5 of 7 observers whose motion perception was influenced by rectangle orientation in Image 2 was 125.6 ± 6.2 msec. Two observers of 7 tested were unable to see any form-based bias in Experiment 3C, and they responded at or near chance for all tested durations. These observers did show an effect in Experiment 3A, although one of them was the observer who was at chance in Experiment 3B. In debriefing, these observers reported that they just saw lines moving between the initial and final images. Their inability to see a form-induced bias in the perceived direction of motion may therefore have been caused by a failure to process the outlines in Image 2 "globally" as rectangles.

Experiment 3D. The Image 2 duration that defined the 75% threshold for the 5 of 7 observers whose motion perception was influenced by the form of Image 2 was 87.7 ± 3.2 msec. The same two observers who could not see coherent motion in Experiment 3C could not see coherent motion here.

Discussion

Experiment 1 established that correspondence matching did not take place on the basis of color. Experiment 3 established that it takes place on the basis of form. In particular, when Image 2 consisted of horizontal rectangles presented for longer than about 100 msec, horizontal transformational motion was perceived, and when it consisted of vertical rectangles, vertical transformational motion was perceived, regardless of color relationships. For example, in Experiment 3A, a red square in Image 1 would transform into a green or red rectangle that overlapped it

in Image 2, which would in turn transform into a red or green square in Image 3 that overlapped it. The 75% threshold can be taken as an estimate of the duration that is needed for a stage of form analysis to complete its analysis, so that correspondence matching can take place on the basis of the analyzed forms. Because form can be defined by color, color does play a role in correspondence matching, even if indirectly. Form analysis functions at and away from equiluminance, and even over line drawings, suggesting that form analysis is primarily concerned with form as defined by contours. If this is correct, then matching will be based on contour-defined form analysis, regardless of whether those contours are defined by color, texture, or some other cue.

The central finding of Experiment 3 is that the stage of form analysis that biases motion processing in TAM takes between 80 and 130 msec. This is significantly slower than the ~50 msec reported by some researchers (Beck, 1982; Julesz, 1984a, 1984b) for texture segmentation. This difference may arise because form analysis may require the segmentation of a scene into figures that are matched across images, whereas texture segmentation only requires segmentation into regions within a single image. Note that this duration approximately coincides with the duration separating element from group motion in the Ternus display (Pantle & Petersik, 1980; Pantle & Picciano, 1976; Petersik & Pantle, 1979) as well as that separating biologically plausible from biologically implausible apparent motion (Shiffrar & Freyd, 1990, 1993). Thus, the duration required to analyze the shape of a new figure and match it to a figure from the previous scene in TAM is approximately the same as the duration required to maintain an existing global perceptual organization over a blank image in translational apparent motion. This does not imply that a common stage of form analysis is involved in translational apparent motion and TAM. The minimal ISI required to see form-based matching in translational apparent motion may have less to do with the duration of form processing than the time required to eliminate biases due to the low-level motion system.

Experiment 3B showed that form analysis at equiluminance takes approximately the same amount of time as under nonequiluminant conditions. TAM, unlike translational apparent motion (e.g., Cavanagh et al., 1985; Ramachandran et al., 1998), is not weakened under conditions of equiluminance. Whereas matching in translational apparent motion may depend primarily on motion energy, luminance contrast may be less important in TAM because matching here takes place on the basis of form relationships. Although the motion processing stream is generally thought to be primarily driven by changes in luminance contrast (e.g., Zeki, 1974), cells in MT respond to moving contours defined by equiluminant color and texture boundaries (Albright, 1992; Dobkins & Albright, 1994; Saito et al., 1989). Such cells may play a role in processing TAM under conditions at and away from equiluminance.

The 75% threshold of Experiment 3C was almost identical to that found in Experiments 3A and 3B. Another experiment, not reported here, revealed that squares and rectangles can be defined by equiluminant texture as well, with similar results (Tse & Logothetis, 2001). It is therefore likely that the form processor is particularly sensitive to contour relationships, regardless of how these contours are defined, whether by color, texture, or outlines. Indeed, we have noted that TAM even occurs over objects defined by illusory contours (Tse et al., 1998).

The 75% threshold found in Experiment 3D was much shorter than any found in the previous experiments and almost 40 msec faster than the result found in Experiment 3C, which was the same experiment except for a reverse of stimulus polarity. Why white outline forms should be more easily or rapidly processed by a stage of form analysis than black outline forms or solid color forms is unclear. A black contour against a white background may be less salient than a white contour against a black background for a number of reasons, including white background glare or higher sensitivity to light increments than decrements. To our knowledge, no extant theory can account for the "white-contour advantage" that we observed in form processing in 5 observers.

EXPERIMENT 4

Experiment 3 established the approximate duration of form analysis involved in biasing TAM. However, it did not address what type of form information is analyzed. The results of Experiment 3 are consistent with a stage of form analysis that operates over contours by, for example, checking the local collinearity of contours. The results, however, are also consistent with a stage of form analysis that takes 2-D figures, such as rectangles and squares, as its primary input. They are also consistent with a stage of form analysis that operates over 3-D representations. In order to better understand the type of information that the stage of form analysis operates over, a new stimulus, derived from the well-known Necker cube, was used. If it can be shown that TAM can be biased toward a 3-D path by 3-D information within the same 80–130 msec time window revealed by Experiment 3, we will be able to conclude that this stage of form analysis is not merely analyzing local contour continuity or 2-D shapes. We will be able to conclude that the stage of form analysis is processing 3-D form information.

If the motion processing system were only taking the 2-D shape or position of successive stimuli as its input, motion along the plane of the screen would be perceived, because this is the plane that contains all TAM stimuli. However, if the motion processing stream is influenced by the inferred 3-D shape of the stimuli, then a smooth shape change could be perceived either out of or into the plane of the screen. If a 3-D motion path is perceived, then we can conclude that the motion processing system has calculated a motion path in 3-D coordinates based on the 3-D forms it derives from 2-D input. To test this hypothesis, subsequent experiments used 3-D stimuli modeled on the Necker cube.

Method

Stimulus presentation was synchronized to the monitor's refresh rate (85Hz). The background was set to the gray value .50, where black is .00 and white is 1.00 on a linear scale. The Necker stimuli were constructed from white lines that were 1 pixel wide. The short Necker stimuli were 10.5° × 10.5°. The long Necker stimulus was 12° × 14°. They were placed so that their rightmost corner was separated from the fixation point by 1 pixel for all images. The fixation point was itself a 0.1° × 0.1° yellow square that was on the screen for the duration of the experiment. The fixation point was not placed directly on the Necker stimulus because doing so made maintaining fixation more difficult. It was placed to the right of the center of the screen so that the Necker stimuli would occupy the center of the screen.

In two experiments (Experiments 4A and 4B), 3 naive observers from the Max Planck subject pool were asked to initiate the transition from Image 1 to Image 2 (see Figure 5) when they were perceiving either one or the other 3-D interpretation of the Necker stimulus in Image 1. Each observer sat 57 cm from the presentation monitor and had the chin in a chinrest. In Experiment 4A, Image 1 was continually present on the screen until the observers pressed a button indicating that they had one or the other of the two 3-D interpretations of the ambiguous Necker stimulus in view. When the observers pressed the button, the short Necker stimulus was instantaneously replaced by the long overlapping Necker stimulus, which then remained on the screen. The observers were given a sheet of paper with the orientations of the stimulus they were supposed to see on each trial. There were 100 trials, 50 for each orientation of the short Necker stimulus. The observers pressed the button when they could see the specified orientation of the stimulus. They then indicated on a piece of paper which orientation of the stimulus they had had in view and whether the motion came forward or went backward or seemed to stay within the plane of the screen, indicating 2-D motion. The observers initiated another trial by again pressing the button, which then returned the short Necker stimulus to the screen. Experiment 4B was similar to Experiment 4A except that, when the observers pressed the button, the short Necker stimulus was replaced by the long Necker stimulus for only 212 msec and then replaced again by the original short Necker stimulus. This created the impression of the short Necker stimulus shooting quickly either forward or backward and then retracting smoothly to its original shape.

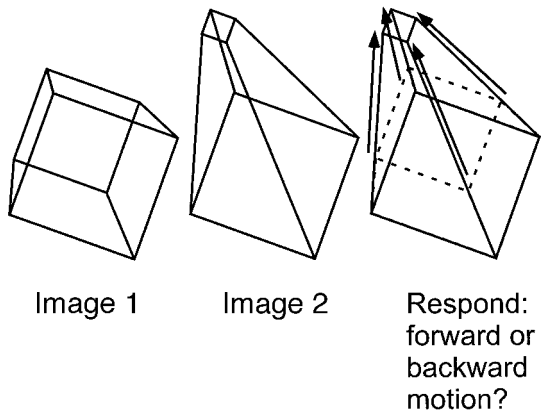

Figure 5. TAM in 3-D. When Image 2 replaces Image 1 all at once, the illusory transformation shown on the right is perceived. The dotted line represents the relative location of Image 1 with respect to Image 2. The Image 1 stimulus appears to grow smoothly either toward or away from the observer, depending on the orientation perceived at the end of Image 1.

Results and Discussion

Results of Experiments 4A and 4B reveal that the perceived direction of elongation was almost always consistent with the 3-D interpretation that existed over Image 1 ($n = 3$; trials = 100; > 95% in consistent direction for both experiments). If the path were merely a cognitive inference of the path that must have been traversed given beginning and end states X and Y, then we would not expect a 3-D motion path to be perceived. A cognitive inference occurs after perception, such as when one infers that it must have rained when the ground is wet. Although perceptual processing itself involves inferential processing (Knill & Richards, 1996), such inferences precede and underlie perception. In particular, TAM is biased by 3-D forms that are constructed prior to the perception of motion. We conclude that the perceived direction of 3-D motion is biased by the form perceived just prior to the switch from the short to the long Necker stimulus. This result is consistent with past results that motion perception operates over internal 3-D models of object shape (Dosher, Sperling, & Wurst, 1986; Sinha & Poggio, 1996).

EXPERIMENT 5

Experiment 4 showed that the perception of TAM follows an analysis of 3-D form. However, in Experiments 4A and 4B, the observers used a top-down strategy to place the ambiguous short Necker stimulus in one 3-D orientation or another. For example, during debriefing, some observers reported that they could choose to see one or the other motion at will. In Experiment 5, the perceived orientation of the ambiguous short Necker stimulus was biased in a stimulus-driven manner, in order to show that perceived 3-D motion in TAM is not merely due to top-down effects.

Method

Observers. The observers ($n = 10$) were instructed to maintain fixation at all times. Observers were obtained from the Max Planck Institute subject pool and were paid. They sat at a distance of 57 cm for all experiments. All observers were naive as to the purpose of the experiment but had taken part in psychophysics experiments before.

Stimuli. The background, fixation point, short, and long Necker stimuli were identical to those used in Experiment 4. The solid biasing shapes shown in Figures 6 and 7 were drawn to convey the impression of a convex shape, using three gray values .98, .70, and .30.

The first image (see Figure 6A) was on for 506 msec and indicated the start of a new trial. The second image was used to bias the 3-D interpretation of the ambiguous Necker stimulus. The experiment consisted of 280 trials: 40 trials at each Image 2 duration (24, 82, 141, 212, 271, 329, and 388 msec; actual durations were next nearest multiples of the 11.76-msec image duration imposed by the monitor's refresh rate of 85 Hz). There were 20 trials for the "forward" orientation of the solid figure and 20 for the "backward" orientation at each duration. Both the duration and the orientation of a solid stimulus were randomized. The third image replaced and exactly overlapped the solid figure instantaneously and remained on the screen for 306 msec, before being replaced all at once by the fourth image for 306 msec. Image 4 was then masked.

Mask. The mask was a sequence of large and small grid patterns composed of white lines of the same width as those that made up

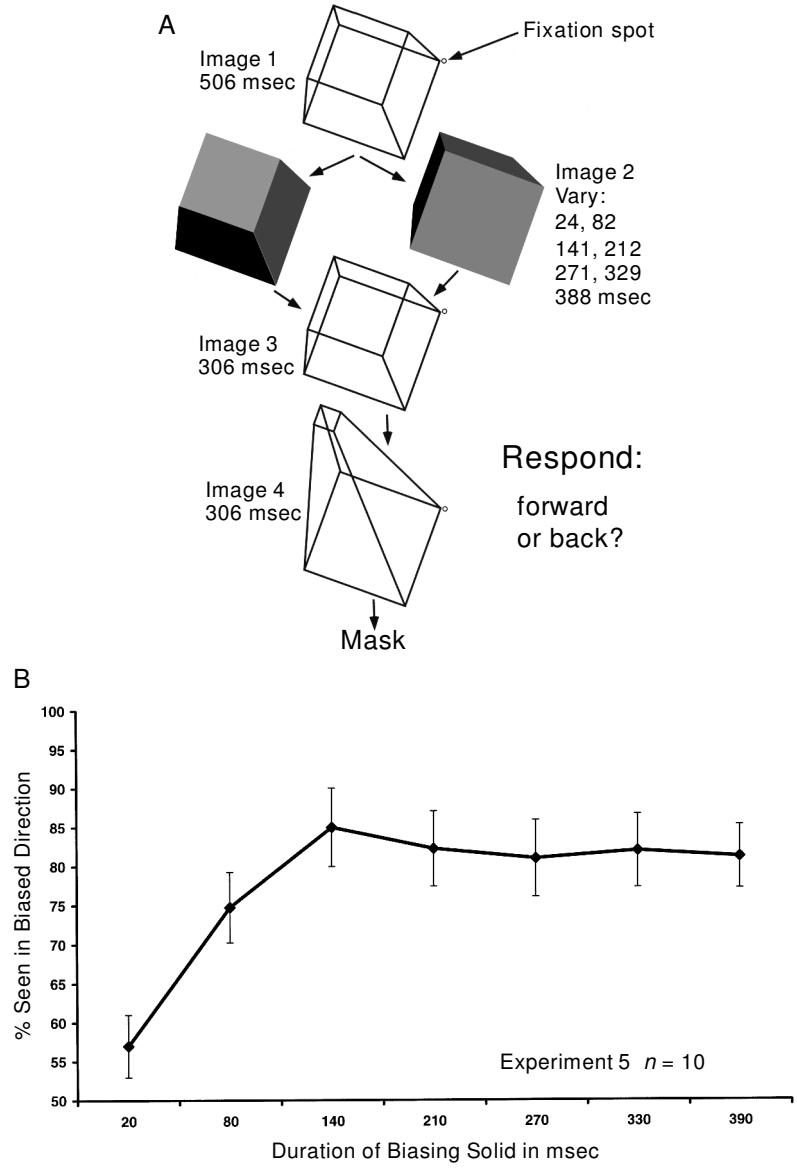

Figure 6. Experiment 5. (A) A Necker stimulus can be biased into one of its two possible 3-D orientations by briefly showing a "solid" figure in one or the other unambiguous orientations. When Image 3 is instantaneously replaced by Image 4, the short Necker stimulus appears to smoothly elongate either toward or away from the observer, depending on the orientation that existed over the Image 3 stimulus just before the transition to Image 4. The duration of the solid stimulus in Image 2 required to bias motion in one or the other direction was measured. Image durations are indicated in the figure. (B) The average percentage of motion seen over Image 4 consistent with the bias imposed by the solid shown in Image 2. Three-dimensional form analysis in the motion pathway can be inferred to take 80–140 msec because the solid must be visible within this range of durations before there is a consistent bias in the direction of perceived motion.

the Necker stimulus. The large grid completely covered the locations occupied by the long Necker stimulus. A slightly smaller grid was also used to generate a moving mask. The large mask was followed by the small mask, followed again by the large and small masks, each for 106 msec, for a total of 424 msec. This mask was then followed by a single image in which the entire screen flashed white for 12 msec. This was done to alert the observer to the end of a trial. The observers had to respond after the flash. The next trial

did not appear until after the observer responded with a buttonpress whether the motion perceived went forward or backward. The intertrial duration was 1,000 msec after the buttonpress.

Results and Discussion

Image 3 of Experiment 5 consisted of a truncated Necker "pyramid," and Image 4 consisted of an elongated version of the same Necker stimulus, as shown in Figure 6A. When Image 3 was replaced by Image 4 all at once without any ISI, such that the Image 4 stimulus exactly overlapped that of Image 3, the Image 3 stimulus appeared to smoothly elongate very rapidly. Indeed, the Image 3 stimulus appeared to elongate either toward or away from the observer along a straight path. If the small end of the Image 3 stimulus appeared to face forward (backward), the direction of elongation was to the front (back). Although many conceivable shape transformation paths could have begun with Image 3 and ended with Image 4, only these two "mini-

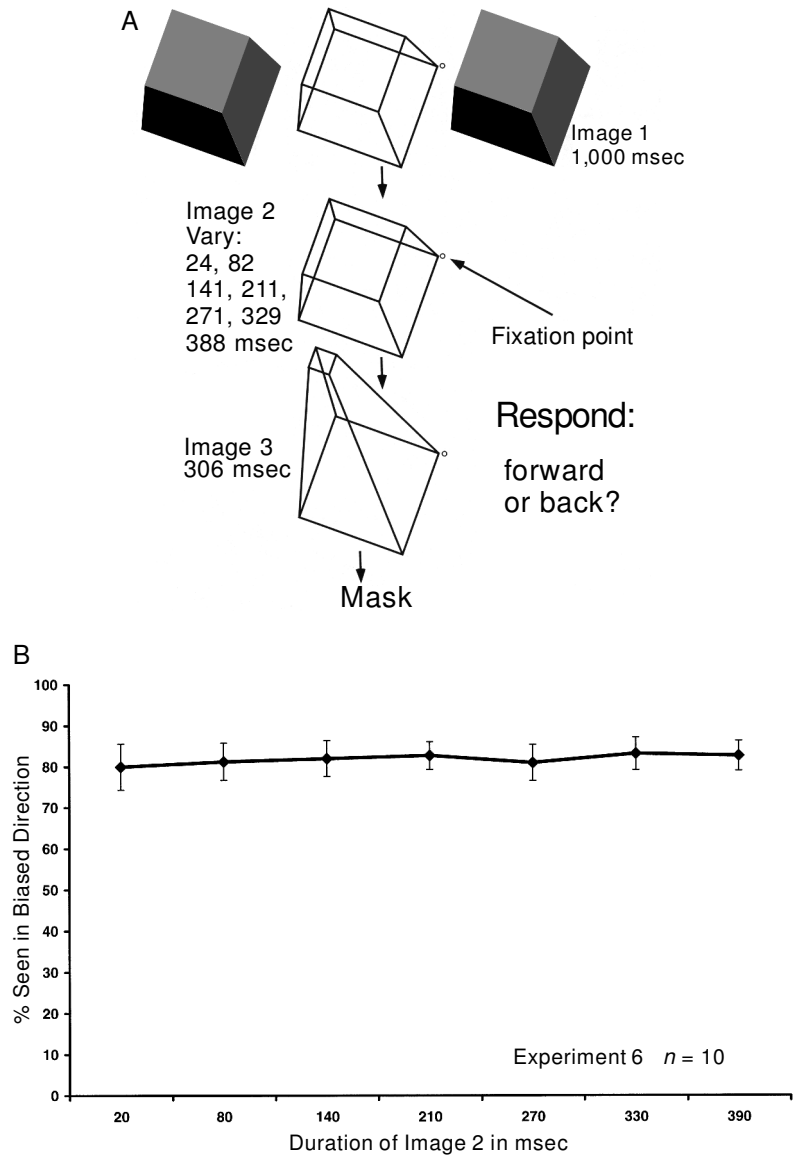

Figure 7. Experiment 6. (A) Two biasing solids tend to "capture" the orientation of the ambiguous Necker stimulus located between them in Image 1. In this case, there is no transient at the location of the Necker stimulus at Image 2, unlike in Experiment 1. There is also no afterimage overlying the Necker stimulus in Image 2. Any bias in the direction of motion is therefore not due to either a transient or an afterimage at the location of the Necker stimulus. (B) The observers saw motion in the direction induced by the flanking solids about 80% of the time up to 380 msec after the flankers disappeared.

mal" transformations were ever reported under these conditions. The direction of TAM could be biased by inserting an unambiguous Image 2 stimulus, establishing that 3-D forms, not just 2-D forms, are analyzed by the stage of form analysis that biases motion perception. Indeed, the motion pathway appears to calculate a path in 3-D that is consistent with the initial and final 3-D forms. The biasing stimulus in Image 2 was a drawing of a "solid" Necker cube in one of its two 3-D interpretations that exactly overlapped the Necker cube of Image 3. The duration of Image 2 that was required to bias the direction of perceived motion was taken to be a measure of the duration required for 3-D form analysis in the motion pathway.

The biasing solid had to be visible for between 80 and 140 msec, in order to bias the 3-D interpretation of Image 3. This duration was consistent with that found in Experiment 3, suggesting that a common form analysis process analyzed both the 2-D forms of Experiment 3 and the 3-D forms of Experiments 4–6. The average ($n = 10$) percentage of motion perceived in the biased direction for Experiment 5 is shown in Figure 6B as a function of the duration of the solid biasing cube in Image 1. Performance reached an average maximum of 85% (motion perceived in the biased direction) at 140 msec and was effectively near chance at 20 msec. Peak performance for some observers was close to 100%; however, for 1 observer, it was as low as 70%. All observers, however, exhibited the same trend in their data. The presentation of Image 3 was kept brief (300 msec) because we found that the Necker stimulus tended to flip automatically for some observers for longer durations, and our goal was to measure the effects of orientation biasing introduced by Image 2 before any spontaneous flips of orientation.

EXPERIMENT 6

We also conducted two control experiments to clarify the results of Experiment 5. Experiment 6 addressed the possibility that Image 2 (see Figure 6A) biased Image 4 directly, by, for example, leaving an afterimage that lingered over Images 3 and 4. On this account, the elongated Necker stimulus did not have the 3-D orientation it did because of a transformation that maintained the orientation of Image 3; rather, it had the same orientation because of the overlaid afterimage. To address this concern, in Experiment 6, we biased the 3-D interpretation of the short Necker stimulus with solid shapes that did not overlap it, and, therefore, did not leave any afterimages over it, as shown in Figure 7A. This also allowed us to eliminate the transient that occurred over the Necker stimulus in Experiment 5 when the solid biasing stimulus was replaced by the short Necker stimulus. The biasing solids in Experiment 6 tended to "capture" the 3-D shape interpretation of the Necker stimulus, even though they did not overlap that stimulus. The two biasing solids were present for 1,000 msec, and, at a variable time after their offset, the short Necker stimulus was instantaneously replaced by the long one.

Results and Discussion

Perceived motion went in the direction of the biasing solids more than 80% of the time for all Image 2 durations tested, as shown in Figure 7B.

Since the biasing solids did not overlap the Necker stimuli in this experiment, the TAM perceived was caused by the 3-D form interpretation of Image 2 and not by any afterimages in Image 2 or transients that might overlap the Necker stimulus.

EXPERIMENT 7

The fact that the observers perceived a forward or backward motion in Experiments 5 and 6 reveals that the motion processing system has calculated a motion path in 3-D as opposed to 2-D coordinates. Despite the compelling illusion of 3-D motion, Experiment 7 addressed the concern that the short Necker stimulus did not transform into the long one in TAM at all. This experiment addressed the possibility that the 3-D interpretation of the long Necker stimulus was merely biased by the 3-D interpretation of the short Necker stimulus. In this control experiment, a single Necker stimulus translated to the opposite side of the fixation spot with a variable ISI, as shown in Figure 8A. The observers had to respond whether the stimulus maintained its 3-D interpretation or was in the opposite 3-D interpretation on translation. If there was a bias to carry 3-D interpretations of the Necker stimulus across images in the absence of TAM, then we could not argue that Image 3 was given its 3-D interpretation in Experiments 5 and 6 because it was a transformation of the 3-D interpretation existing over the short Necker stimulus. However, if there were no tendency to carry 3-D interpretations across successive images in the absence of TAM, then it is likely that the 3-D interpretation existing over the short Necker stimulus actually transforms into that of the long Necker stimulus.

Method

The same short Necker stimulus used in Experiments 4–6 was used here. It had the same placement with respect to the fixation point. It was presented on either the left or right side of the fixation point for 1,000 msec, followed by presentation on the opposite side for 1,000 msec, followed by a mask of the same type used in Experiments 5 and 6 whose width and height covered the positions of both Necker stimuli. The two presentations of the Necker stimulus were separated by a blank ISI of 24, 71, 141, 212, 271, 329, or 388 msec. There were 280 trials. In 140 trials, the initial position was on the left (right) side of fixation. Initial position was randomized. The observers ($n = 6$) were also the observers in Experiments 5 and 6. It was emphasized that even if the Necker stimulus spontaneously flipped during the initial or final states of a given trial, the important orientations for comparison were the last orientation they saw for the initial Necker stimulus and the first orientation they saw for the final Necker stimulus. The observers pressed one of two buttons to indicate whether these orientations were the same or different.

Results and Discussion

The percentage of trials in which a Necker stimulus appeared to translate into a second Necker stimulus and maintain the same 3-D orientation was on average at

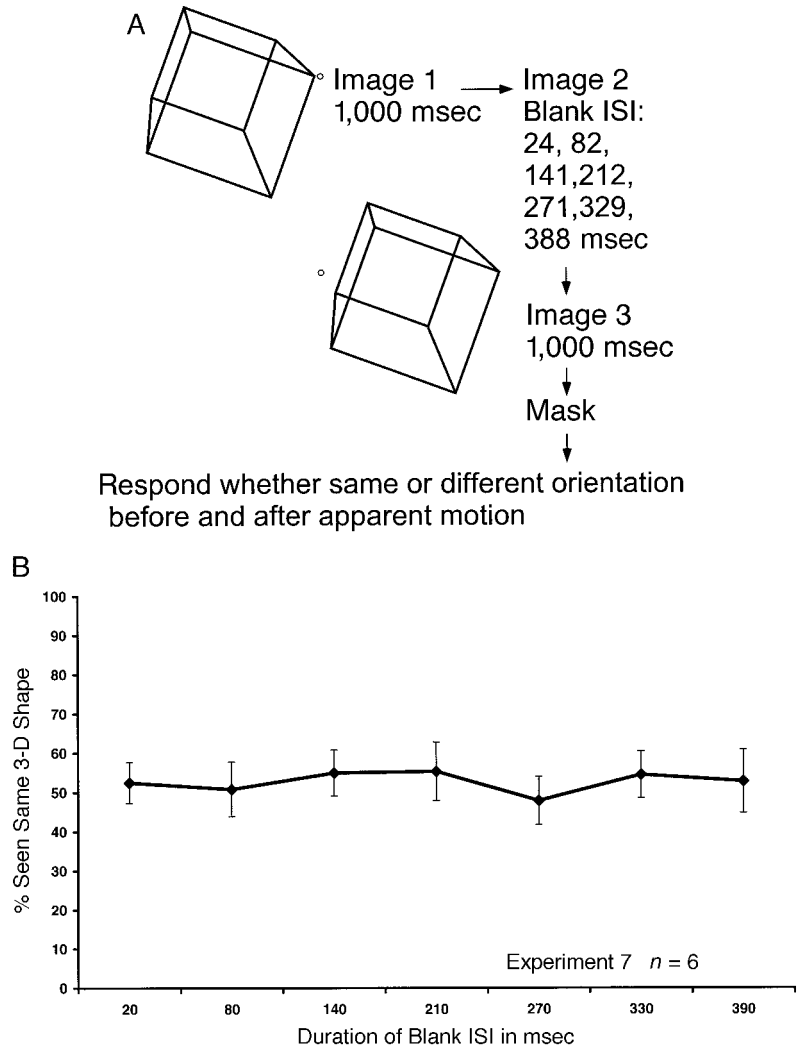

Figure 8. Experiment 7. (A) The Necker stimulus was shown in Image 1 to the left or right of fixation for 1,000 msec, followed by a variable blank interval between 20 and 380 msec, and presentation of the same Necker stimulus on the other side of fixation for 1,000 msec, followed by a mask. The observers had to respond whether the Necker stimulus had the same orientation before and after it underwent apparent motion to its new position. (B) The observers, on average, reported the same orientation before and after apparent motion as often as they reported the opposite orientation. This control experiment demonstrates that the tendency for the observers to see the same orientation under TAM in Experiments 5 and 6 was not due merely to temporal coincidence. Rather, it must have been due to the spatial relationship of the long to short Necker stimulus, so that the long one was seen as a transformation of the short one.

chance for all ISIs, as shown in Figure 8B. This indicates that the tendency to maintain the Image 3 interpretation as the stimulus changed into Image 4 in Experiment 5 was not simply due to a tendency to interpret Image 4 in the same way as Image 3 but was due to a transformation of the Image 3 configuration into that of Image 4 (see Figure 6A). In sum, Experiment 6 ruled out biasing due to afterimages, and Experiment 7 ruled out the possibility that the initial perceived form merely biases the interpretation of perceived form in the final image, without playing any role in biasing the direction of perceived transformational motion.

GENERAL DISCUSSION

TAM is not the first motion phenomenon to reveal the importance of one or more stages of form analysis in motion processing. Past research on smoothly moving contours has suggested that contour relationships contribute to motion analysis. In particular, Wallach (1935, 1976;

Wuerger, Shapley, & Rubin, 1996) found that the perceived direction of motion of a uniformly colored straight line drifting smoothly behind an aperture depends on the shape of that aperture, even though such a line does not possess a mathematically well-defined direction of motion. In doing so, Wallach presaged what many years later would be called *the aperture problem* (Hildreth, 1984; Marr, 1982; Marr & Ullman, 1981; Movshon, Adelson, Gizzi, & Newsome, 1985). Later authors (Adelson & Movshon, 1982; Burt & Sperling, 1981; A. B. Watson & Ahumada, 1985) argued that the aperture problem could be solved by combining different component motions. Adelson and Movshon, for example, proposed a two-stage process to account for motion integration. A stage in which oriented motion detectors were tuned to the various directions of component motion was followed by a stage of intersection of constraints. This motion energy model did not address the role of terminators on perceived direction and, as such, did not posit any direct role for form in motion processing. Wallach, however, had found that the perceived direction of motion was linked to both the perceptual organization of the scene and the nature and motion of line endpoints. He made the distinction between what would later be called *intrinsic* and *extrinsic* line terminators (Shimojo, Silverman, & Nakayama, 1989; i.e., terminators that arise from endpoints belonging to the moving line, and those arising from points where the moving line is occluded). Recent work has shown how such terminator motions influence processes, such as amodal completion and global integration of local motion signals (Lorenceau & Shiffrar, 1992; Shiffrar, Li, & Lorenceau, 1995).

Although contour terminators play a role in motion perception that cannot be adequately explained in terms of motion energy, it has not been clear that 3-D form per se influences motion perception. Contour terminators could still be functioning as a local cue that influences a motion system dominated by motion energy detecting comparators. However, more is involved in form–motion interactions than local contour terminators. For example, there are several illusions in which a 3-D shape appears to change its direction of rotation depending on the 3-D form interpretation one places over the moving object (e.g., Ames's rotating trapezoidal window illusion, Ames, 1951; the rotating mask illusion, e.g., Klopfer, 1991; motion from structure, e.g., Ullman, 1979). Others (Dosher et al., 1986) have noted that when a Necker stimulus is continuously rotated from a stationary position, it appears to rotate in a direction consistent with the 3-D interpretation it had when stationary. Our Experiments 4A and 4B replicated this finding in the context of TAM. Sinha and Poggio (1996) showed that the representation of the 3-D form of an ambiguous "rotating" wire silhouette determines whether rigid rotation or deformation is seen. They rotated a computer-generated wire silhouette. Although an infinite number of 3-D rotations are consistent with the silhouette motion, an assumption of object rigidity allows the extraction of a single 3-D shape by observers. When a new wire is rotated from an initial position that happens to cast the same silhouette as the first wire, observers tend to see the wire deform, as the silhouette takes on shapes inconsistent with the shape inferred from the first rotating silhouette. Interestingly, observers who do not receive training with the first wire do not see deformation in the second stimulus but instead see rigid rotation. This demonstrates both the existence of an object rigidity assumption and the existence of an internal 3-D model that can bias perceived motions toward paths involving rigid rotation or deformation. These studies have convincingly shown that global form analysis plays a role in the perception of real (i.e., continuous) motion.

Evidence of form–motion interactions using translational apparent motion as a probe is less common. Indeed, most studies using translational apparent motion as a probe have reported little, if any, influence of form on correspondence matching (e.g., Burt & Sperling, 1981; Krumhansl, 1984). However, there are studies that show that the type of apparent motion perceived depends on the ISI between images of a translational apparent motion sequence. For example, in the Ternus display (Ternus, 1926/1938), there are three equidistant, collinear dots at Positions A, B, and C. These are replaced after a variable ISI by three other equidistant, collinear dots at Positions B, C, and D. Observers tend to see two different types of motion. In "element motion," observers see the dot that was at A jump to Position D, whereas the dots at Positions B and C flicker but stay in place. Under "group motion," observers see the dots move as a group from Positions A, B, and C to Positions B, C, and D, such that each dot shifts over one position. Element motion tends to dominate for ISIs shorter than about 100 msec, and group motion tends to dominate for longer ISIs (Pantle & Picciano, 1976). It has been proposed that this results from interaction between a low-level or short-range and high-level or long-range motion processing system (Braddick, 1974, 1980). The low-level system detects whether there has been a positional change over the inner dots at Positions B and C. If no low-level motion is signaled, perhaps due to visual persistence of the overlapping dots (Breitmeyer & Ritter, 1986a, 1986b), the high-level system concludes that the inner dots are flickering but stationary objects. Correspondence must therefore take place between the dots at Positions A and D, leading to element motion. However, when the ISI is longer, or when the positions of the dots at B and C have shifted more than 16′ of arc, group motion is perceived, because the low-level motion processing system has signaled local motion of the dots (Pantle & Petersik, 1980; Pantle & Picciano, 1976; Petersik & Pantle, 1979). The high-level system can then match at the level of the dots taken as a group.

ISI is not the only factor affecting whether element or group motion will be perceived. He and Ooi (1999) provided evidence that another factor determining whether group or element motion is perceived is the perceptual organization of the dots prior to their displacement. When the dots are organized as a group they will tend to undergo group motion, and, when they are not, they will tend to undergo element motion. Note that the results of our

Experiment 7 imply that a 3-D form interpretation is not preserved over large displacements in translational apparent motion. Overlapping stimuli, as in TAM, or small displacements in translational apparent motion, such as the partially overlapping displacement used in the Ternus display by He and Ooi, allow the preservation of 3-D form interpretations despite discrete shape or position changes. In sum, the form that one sees in a given scene will tend to be preserved as the perceived form moves, whether this is continuous motion (Dosher et al., 1986; Sinha & Poggio, 1996), translational apparent motion with small displacements (He & Ooi, 1999), or TAM (Experiments 4–6).

Shiffrar and Freyd (1990, 1993) considered apparent motion over photographs of human bodies and found that observers tended to see biologically plausible motion, such as a hand moving around a body from the back to the front, only when the ISI between photographs was higher than approximately 100 msec. Below this duration, observers tended to see biologically implausible motion, such as a hand passing through a body. This ISI "border" between conflicting apparent motion percepts coincides with that found using the Ternus display, suggesting that a common stage of form analysis may play a role in both types of translational apparent motion display. Biologically plausible apparent motion may be analogous to group motion and biologically implausible apparent motion may be analogous to element motion. An existing global form interpretation can be carried over to the next image when the biases introduced by the low-level motion processing system are minimized by having a long ISI between the images. For short ISIs, the biases of the low-level system dominate, and these appear to be indifferent to considerations of biological plausibility. Instead, the low-level motion processing system appears to match on the basis of motion energy, spatial frequency (Green, 1986b; Ramachandran et al., 1983), and a minimization of path distance (Ullman, 1979).

The results described using the Ternus display or biological apparent motion do not offer any estimate of the duration of the form analysis that can bias motion processing. Nor do these results specify the visual cues analyzed by this stage of form analysis. Results to date merely indicate that the form interpretation of Image A can bias the form interpretation of Image B (including the perceived motion path between A and B), when low-level biases are minimized by an ISI of sufficient duration. To our knowledge, no researchers have offered estimates of the duration of form analysis involved in biasing the perceived motion path in apparent motion displays. The closest to date have been reports of the duration that a blank ISI must have in order to maintain an existing form interpretation over successive images of a translational apparent motion stimulus (He & Ooi, 1999; Pantle & Petersik, 1980; Pantle & Picciano, 1976; Petersik & Pantle, 1979; Shiffrar & Freyd, 1990, 1993). This duration (~100 msec) is similar to the duration of form processing that we have inferred using TAM. This does not prove that translational apparent motion and TAM rely on common mechanisms of form or motion analysis, although this might be so. However, the minimum ISI required to see effects of form-based matching in translational apparent motion may have less to do with the duration of form analysis than with the time needed to eliminate biases due to matching on the basis of low-level cues, such as motion energy. TAM, in contrast, does not require a blank ISI between images in order to see the results of form-based matching and may actually be diminished by such an ISI. For these reasons, TAM is a better stimulus for determining characteristics of form analysis in the motion pathway once we control for matching on the basis of motion energy. We believe our results offer the first direct measure of the duration of form analysis in the context of motion processing.

Our experiments have revealed two properties of the stage of form analysis that biases the shape transformation seen in TAM. First, a biasing stimulus must be present for processing for 80–140 msec in order to bias the motion path. This suggests that form analysis takes at most this long. If form analysis took longer, a bias could not arise this quickly. If form analysis took substantially less time, we would expect shorter presentation times to effectively bias the motion outcome. Second, the biasing stimulus can be either 2-D or 3-D, and 3-D form analysis does not take longer than 2-D. This implies that this stage of form analysis is organized to take 3-D form as its input. Because contours can be defined by multiple cues, including outlines, it is probable that the stage of form analysis analyzes contour relationships in order to extract 3-D form. The motion that is perceived is a continuous transformation that maintains the 3-D interpretation that existed over the stimulus just before its transformation. Indeed, the transformation that is perceived is the minimum transformation that could have changed the object perceived in the initial image into the object perceived in the final image. In obeying this "minimal transformation assumption," the visual system interpolates the path that involves the least motion in 3-D between the initial and final states. Others have suggested similar motion minimization rules (Ullman, 1979; Wertheimer, 1912/1961).

Experiments 1, 3A, and 3B provided no evidence for a direct role of color in matching (contrary to Green, 1986a, 1986b, 1989). Color, however, does play an indirect role in matching figures across images because color can specify figure boundaries, and figures are matched across images. One prediction that emerges from these experiments is that (controlling for motion energy and attentional tracking) TAM will occur, no matter how contours and shape are defined. As long as contour relationships among successive images are ones that the form processor can use to segment one figure from another, it should be possible to match figures defined by different cues. Our experiments imply that the maintenance of contour collinearity is a strong indication that two regions in successive images comprise different states of the same figure. But contour collinearity can be defined differently in successive images. In particular, it should be possible to see TAM even when a figure in the first image of a two-image TAM sequence is defined by one set of texture differences between figure and ground and by another set of texture differences in the sec-

ond image. Indeed, it may be possible to see TAM occur between a 3-D object defined by, for example, shape-from-motion (Ullman, 1979) in one "image" and a transformed 3-D object defined by outlines or shading in the next. What our results imply is that the motion processing system constructs transformational paths over internal 3-D models of objects regardless of the shape-from-*x* information used to construct those models in the first place. We are currently conducting experiments to determine whether these predictions are correct.

This discussion of "form-cue invariant TAM" is reminiscent of a recent debate about second-order motion. Cavanagh et al. (1989) found evidence that translational apparent motion can occur between first-order (luminance-defined) and second-order (contrast-defined) stimuli, suggesting that both types of motion are processed by a common motion detecting mechanism. More recently, several authors (Ledgeway & Smith, 1994, 1995; Mather & West, 1993; Seiffert & Cavanagh, 1998) have provided evidence that first-order motion and second-order motion are not processed by a common motion detector type. Rather, they are processed by separate low-level detectors, each insensitive to motion of the other class. Edwards and Badcock (1995), however, provided evidence that, unlike first-order motion detectors, second-order motion detectors were sensitive to motion of the other class. This would be expected if second-order filters detect or are influenced by contour, texture, or form information, because form can be defined using first-order cues. Wilson, Ferrera, and Yo (1992; see also Derrington, Badcock, & Henning, 1993; Zanker & Huepgens, 1994) suggested that first-order motion is processed foremost in V1, whereas second-order motion also requires processing in V2, with both streams converging in MT. This was supported by the finding that 87% of sampled cells in MT in the alert macaque are tuned to second-order motion as well as first-order motion (Albright, 1992, see also Olavarria, DeYoe, Knierem, Fox, & Van Essen, 1992; but compare O'Keefe & Movshon, 1998, who found $< 25\%$ such cells in MT in the anesthetized macaque; for second-order motion tuning in cat cortex, see Mareschal & Baker, 1998, 1999, and Zhou & Baker, 1993, 1994, 1996). Albright (1992) suggested that these cells may underlie form-cue invariant motion processing (see also Buracas & Albright, 1996). Form-cue invariant TAM, and TAM more generally, may be processed in MT (compare Kawamoto, Yoshino, Suzuki, & Ichinowatari, 1997) by such cells and may have much in common with second-order motion.

Several authors have argued that second-order motion processing requires a stage of texture extraction (Derrington & Henning, 1993; Stoner & Albright, 1993; Werkhoven, Sperling, & Chubb, 1993). This stage of texture extraction (presumably carried out in V2, V3, and/or V4; compare Smith, Greenlee, Singh, Kraemer, & Hennig, 1998) may correspond to the stage of form analysis that we argue must precede the assignment of motion paths in TAM. Of course, there need not be a single form processor underlying TAM, just as there need not be a single stage of form extraction underlying second-order motion. There may exist several shape-from-*x* systems, just as there may be multiple types of second-order motion detectors (Petersik, 1995). But for form-cue invariant TAM to be possible, the various shape-from-*x* systems and/or second-order motion detectors must converge on a common representation of moving shape. Texture features, once extracted, could correspond to the figures that we argue are transformed in TAM. Given the results of Experiment 5, we would argue that the stage of texture extraction in second-order motion is actually a stage of 3-D form analysis and that the reason earlier authors hypothesized a stage of texture as opposed to 3-D form extraction was that their stimuli were texture-defined gratings rather than 3-D objects.

In general, TAM is influenced not only by form relationships among stimuli but also by attentional tracking and by motion energy. Here, we have controlled for these other two factors in order to study the properties of the form extraction process. Similarly, motion among extracted second-order features or figures is influenced by attentional or position tracking, coinciding with evidence that attention can bias TAM (e.g., Hikosaka et al., 1993a, 1993b). Although Lu and Sperling (1995) argued that neither first- nor second-order motion depends on tracking, there is evidence that, at low contrasts, second-order motion is dependent on tracking (Derrington & Ukkonen, 1999; Seiffert & Cavanagh, 1999).

Baloch and Grossberg (1997) modeled TAM to involve three interacting subprocesses: (1) a boundary completion process (V1 $\rightarrow$ interstripe V2 $\rightarrow$ V4); (2) a surface filling-in process (blob V1 $\rightarrow$ thin stripe V2 $\rightarrow$ V4); and (3) a long-range apparent motion process (V1 $\rightarrow$ MT $\rightarrow$ MST). These three processes have been described in earlier work by Grossberg and colleagues (Cohen & Grossberg, 1984; Francis & Grossberg, 1996a, 1996b; Grossberg, 1994; Grossberg & Mingolla, 1985a, 1985b; Grossberg & Rudd, 1992; Grossberg & Todorović, 1988), but, in order to accommodate TAM, Baloch and Grossberg added a new link between V2 and MT in their model to allow the motion processing stream to track emerging boundaries and filled-in surface colors. According to the model, these three processing streams generate three separate moving waves of cortical activation—(1) a boundary completion wave, (2) a wave of filled-in color, and (3) a wave of motion energy—any or all of which might underlie the percept of TAM. The model is foremost a bottom-up account of boundary completion, where existing boundaries enhance growth of neighboring collinear boundaries and inhibit growth of neighboring dissimilarly oriented boundaries. Color is then filled in from existing color regions to new color boundaries. The model accounts well for the fact that TAM tends to proceed away from the cue with which it shares collinear contours, as in Figures 1 and 2. The model also agrees with our view that TAM emerges from mechanisms dedicated to constructing, filling in, and tracking 3-D shapes over time.

There are, however, problems with the Baloch and Grossberg (1997) model. One implication of the model is that the three waves can go in opposite directions. Pre-

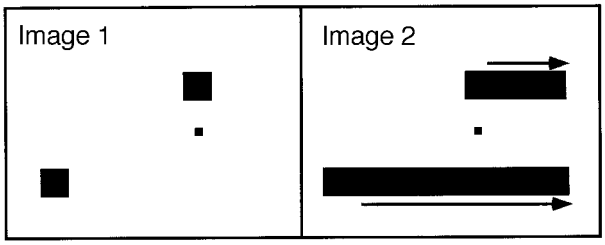

Figure 9. When two spots undergo TAM into lines at the same time, but one line is longer than the other, TAM appears to complete for both lines at the same time. The speed of TAM in the longer line appears to be faster than that for the shorter line. The small spot indicates the fixation point.

sumably, the boundary and filling-in waves would proceed from left to right in Image 2 of Figure 2, but the motion wave would proceed in the opposite direction (because the centroid of the new parts of Image 2 are closer to the right cue). Why certain waves should be perceived and others not is not clear from their model. Also, the mechanism that subserves propagation of their boundary completion wave is one of lateral excitation/inhibition among neighboring cells tuned to boundaries. Because this is a local process, the rate of boundary completion should be a constant number of degrees per second at a given location in the visual field dependent solely on the speed with which neighboring cells can excite/inhibit one another. However, in an example such as the one shown in Figure 9, the longer and shorter lines in Image 2 appear to complete at the same time, and the speed of the longer line as it undergoes TAM appears to be faster than that of the shorter line. This argues against a constant rate of boundary completion or of color filling-in. Another difficulty for the model emerges from the local nature of the boundary interactions they describe. TAM is influenced by global configural relationships among stimuli. For example, when the stimulus configuration shown in Figure 10, Image 1, is replaced all at once by Image 2a, motion inward from both the left-hand cue and the right-hand cue is perceived. However, if Image 1 is replaced all at once by Image 2b instead of Image 2a (Tse et al., 1996, 1998; see also Downing & Treisman, 1997), then the percept between the cues is very different. Now the line between the cues undergoes TAM entirely away from the left-hand cue. This means that the mechanism underlying TAM must be sensitive to global configural relationships within and between the two images. Purely local lateral excitation and inhibition are not sufficient to account for TAM.

Although aspects of Baloch and Grossberg's (1997) model are reasonable, and their boundary completion process carries out much of what a stage of form analysis would have to accomplish in order to separate overlapping figures from one another on the basis of contour relationships, we believe that the stage of form analysis must also instantiate a more global analysis of figural relationships between and within successive images. In particular, the visual system seems to operate under two ecologically reasonable and related assumptions. One is that objects do not vanish or appear out of thin air (Tse et al., 1998). Rather, they continuously change into or occlude one another across successive images. The other assumption is that objects change in such a way that each object is mapped onto at least one object in the next image, and the number of motion paths are minimized under this constraint. This is similar to Ullman's (1979) notion of "minimal mapping," which he described as a solution to the correspondence problem for translational apparent motion. Although it would be desirable to have a global minimal mapping fall out of simple local interactions, no model of TAM to date can account for results such as those shown in Figure 10. It may be that a minimal mapping principle is realized in terms of local interactions that have not yet been modeled. However, it is also possible that minimal mapping requires global processing that cannot be reduced to lateral excitation and inhibition among neighboring cells in early visual areas. Such global processing may involve operations over more abstract classes of information than the kind that single cells in early visual areas are commonly thought to be tuned to. For example, if the stage of form analysis is concerned with matching each figure in Image 1 to at least one figure in Image 2, even if this requires the construction of motion paths that traverse several degrees or even tens of degrees, then explanations of TAM in terms of purely local interactions among neurons tuned only to image properties such as edges or color will be difficult to maintain. Rather, future models of TAM will most likely not be purely local in nature and will have

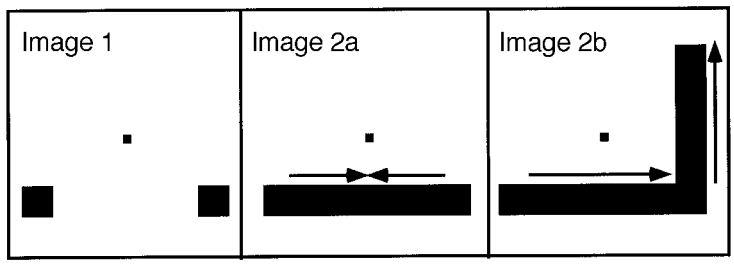

Figure 10. When Image 1 is replaced all at once by Image 2a, inward TAM is perceived away from both cues, as indicated by the arrows. However, when Image 2b replaces Image 1, the line between the two cues undergoes TAM entirely away from the left cue. The small spot indicates the fixation point.

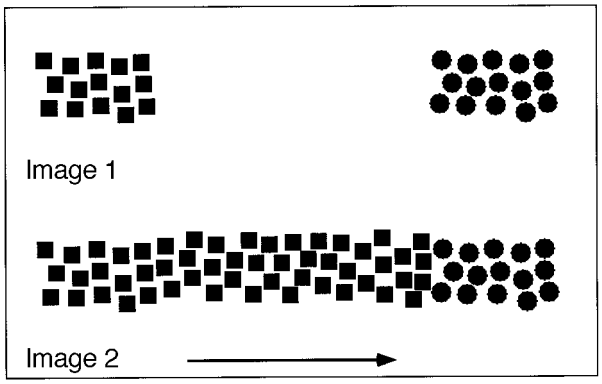

Figure 11. TAM proceeds away from a cue that shares a common texture, even when there are no explicit contours defining the cues. The arrow indicates the most common percept of motion.

to incorporate internal operations involving figure tracking (compare Ullman's [1988, 1995] notion of "visual routines" or Pylyshyn's [1989] notion of "FINSTs").

We have said that the stage of form analysis that influences TAM takes contours as its input. This would be too stringent a conclusion, if by contours we meant only high spatial frequency object boundaries. Object boundaries are likely to be constructed at multiple spatial scales because the visual scene is sampled at multiple spatial scales by neurons with different receptive field sizes. Therefore, an object that has fuzzy or ambiguous boundaries should also undergo TAM. The job of the stage of form analysis that influences TAM is presumably to define spatiotemporally coherent objects. What counts as an object is an open question. Objects may appear distinct, like a cube, or they may appear fuzzy, like a swirl of fog. If this stage of form analysis is also involved in grouping, TAM may occur over elements that have been grouped, though not necessarily grouped into coherent objects with distinct contours. For example, in Figure 11, TAM originates from the cue that would be predicted on the basis of grouping procedures. However, controlled studies are needed to determine whether TAM is actually originating in such cases on the basis of grouping cues or on the basis of some other cue, such as differing luminance contrast or spatial frequency information. Indeed, TAM can be used as a probe to explore what counts as an object within the motion processing stream.

The fact that a 3-D motion path is seen, as in Experiments 4–6, means that the motion processing system is organized to calculate and track motion in a 3-D coordinate frame. Although the results of Experiments 1–3 could have arisen given a form processor that analyzed only local contours or 2-D shapes, the results of Experiments 4–6 imply that, even in Experiments 1–3, 3-D form and motion were processed. That is, the observed 2-D motions in Experiments 1–3 do not arise because the form analyzer can process only 2-D forms or because the motion processor can calculate only 2-D paths. Two-dimensional motions were observed in Experiments 1–3 because these were the motion paths dictated by the stimulus configurations tested.

It may seem surprising that form analysis takes the same amount of time whether the stimuli are 2-D (80–130 msec in Experiment 3) or 3-D (80–140 msec in Experiment 5). Structural theories of form perception and object recognition (e.g., Leeuwenberg, 1969) might suggest that a cube would take longer to perceive than a square, because there is more shape information to process in a cube than in a square. However, it would be pointless to have a 2-D shape analysis process for 2-D shapes and to have a 3-D one for 3-D shapes, because the stage of form analysis cannot know in advance if it is processing 2-D or 3-D shapes. Because all shapes in the world are 3-D, it is not surprising that all shapes are processed for their 3-D form. If the form processor analyzes all forms as 3-D, then it may not take less time to analyze the form of a square than the form of a cube. A square would then just be a flat 3-D object constructed by a form analysis process that takes a fixed minimal duration to calculate 3-D form.

It is conceivable that form analysis would operate on each image independently, followed by an independent stage of matching based on form correspondence among elements in successive images. However, form analysis and correspondence matching are not independent or serial processes. Because the form interpretation reached for a given figure takes into account the form of figures in the previous scene, form analysis and matching are part and parcel of the same spatiotemporal process. Matching occurs because figures are defined spatiotemporally, rather than just spatially in each successive scene. In other words, matching is subsumed by a figure formation process that operates over a certain range of spatiotemporal extents, and what gets matched across scenes are corresponding 3-D figures, where correspondence is determined by the assumption of minimal transformation in a 3-D sense. Thus, form as processed by the visual system is spatiotemporal rather than just spatial, and the dichotomy between (spatial) form or figure analysis and the (temporal) matching of figures across images is a false and misleading one (recent work emphasizing the importance of spatiotemporal form and motion processing can be found in Gepshtein & Kubovy, 2000, Tse et al., 1998, and Wallis & Bülthoff, 2001; the first proponent of an entirely spatiotemporal analysis of form and motion may have been Gibson, 1979).

When the transition from initial to final states is too rapid to be detected by the visual system, it appears that the visual system "fills in" the 3-D transformational motion path that must have been traversed, given the initial and final states and given an assumption of minimal but continuous transformation. Why are intermediate states of an object filled in so much more compellingly in TAM than in translational apparent motion? If one fixates between two spots jumping back and forth in translational apparent motion, the spot does not compellingly appear to occupy intermediate positions, although one nonetheless senses motion between the two positions (Wertheim-

er, 1912/1961). Why does the visual system not fill in the phenomenal features of a moving spot at intermediate positions in the case of translational apparent motion? If a spot had really traversed these intermediate positions, it presumably would have left detectable evidence, such as a blur, of its traversal. But when no such trace is detected, the motion processing system may assume, by default, that the nearest neighbor path has been traversed, whereas processes responsible for filling in phenomenal features may not fill in that for which there is no perceptual evidence. The motion system might interpolate a motion path even in the absence of image data from intermediate positions, because the motion system has to interpolate paths for objects that move behind occluders, such as a bird flying behind leaves. In contrast, a system responsible for phenomenal filling-in of "missing" image data would not fill in the missing data in the absence of other image data supporting the interpretation that intermediate positions had been occupied and were not occluded. Note that when image "evidence" of a path traversal is added, nondefault paths can be perceived. Shepard and Zare (1983) noted that if an artificial blur is added for as short as a few milliseconds, a spot can appear to traverse curved paths. It is not clear from their results whether the spot was seen to continually traverse these intermediate positions. In TAM, the surface itself specifies the path between initial and final states and therefore provides evidence of the path that should be phenomenally filled in. Translational apparent motion and TAM may look different because TAM involves interpolation of a motion path as well as phenomenal filling-in of intermediate traversed states, whereas translational apparent motion may involve only path interpolation with little or no phenomenal filling-in of intermediate states.

In conclusion, forms are interpreted to transform smoothly in TAM, even though inputs are discrete. A stage of 3-D form analysis underlies TAM and may relate the illusion to second-order motion. Like form, transformational motion paths are constructed within a 3-D coordinate frame. The duration of this stage of form analysis is approximately 100 msec.

REFERENCES

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, **2**, 284-299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, **30**, 523-525.

Albright, E. H. (1984). Direction and orientation selectivity of neurons in cortical area MT of the macaque. *Journal of Neurophysiology*, **52**, 1006-1130.

Albright, E. H. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, **255**, 1141-1143.

Ames, A. (1951). Visual perception and the rotating trapezoidal window. *Psychological Monographs*, **65**(7).

Anstis, S. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London: Series B*, **290**, 153-168.

Anstis, S., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In J. D. Mollon & L. T. Sharpe (Eds.), *Color vision: Physiology and psychophysics* (pp. 156-166). New York: Academic Press.

Baizer, J., Ungerleider, L., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience*, **11**, 168-190.

Baloch A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formotion dynamics. *Vision Research*, **37**, 3037-3059.

Barlow, H., & Levick, W. R. (1965). The mechanism of directionally sensitive units in rabbit's retina. *Journal of Physiology*, **178**, 477-504.

Baro, J. A., & Levinson, E. (1988). Apparent motion can be perceived between patterns with dissimilar spatial frequencies. *Vision Research*, **28**, 1311-1313.

Beck, J. (1982). Texture segregation. In J. Beck (Ed.), *Perceptual organization and representation* (pp. 285-317). Hillsdale, NJ: Erlbaum.

Braddick, O. J. (1974). A short-range process in apparent motion. *Vision Research*, **25**, 839-847.

Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London: Series B*, **290**, 137-151.

Breitmeyer, B. G., & Ritter, A. (1986a). The role of visual pattern persistence in bistable stroboscopic motion. *Vision Research*, **26**, 1801-1806.

Breitmeyer, B. G., & Ritter, A. (1986b). Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception & Psychophysics*, **39**, 275-280.

Bundesen, C., Larsen, A., & Farrell, J. E. (1983). Visual apparent movement: Transformations of size and orientation. *Perception*, **12**, 549-558.

Buracas, G. T., & Albright, T. D. (1996). Contribution of area MT to perception of three-dimensional shape: A computational study. *Vision Research*, **36**, 869-887.

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, **88**, 171-195.

Cavanagh, P., Arguin, M., & von Grünau, M. (1989). Interattribute apparent motion. *Vision Research*, **29**, 1197-1204.

Cavanagh, P., Boeglin, J., & Favreau, O. E. (1985). Perception of motion in equiluminous kinematograms. *Perception*, **14**, 151-162.

Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology*, **71**, 2269-2280.

Cohen, M. A., & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception & Psychophysics*, **36**, 428-456.

Dawson, M. R. W. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, **33**, 569-603.

Derrington, A. M., Badcock, D. R., & Henning, G. B. (1993). Discriminating the direction of second-order motion at short stimulus durations. *Vision Research*, **33**, 1785-1794.

Derrington, A. M., & Henning, G. B. (1993). Detecting and discriminating the direction of motion of luminance and colour gratings. *Vision Research*, **33**, 799-811.

Derrington, A. M., & Ukkonen, O. I. (1999). Second-order motion discrimination by feature-tracking. *Vision Research*, **39**, 1465-1475.

Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, **237**, 400-402.

Dobkins, K. R., & Albright, T. D. (1994). What happens if it changes color when it moves? The nature of chromatic input to macaque visual area MT. *Journal of Neuroscience*, **14**, 4854-4870.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, **26**, 6, 973-990.

Downing, P., & Treisman, A. (1995). The shooting line illusion: Attention or apparent motion? *Investigative Ophthalmology & Visual Science*, **36**, S856.

Downing, P., & Treisman, A. (1997). The line-motion illusion: Attention or impletion? *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 768-779.

Edwards, M., & Badcock, D. R. (1995). Global motion perception: No

interaction between the first- and second-order motion pathways. *Vision Research*, **35**, 2589-2602.

Exner, S. (1875). Über das Sehen von Bewegungen und die Theorie des zusammengezetzen Auges [On motion perception and the theory of compound eyes]. *Sitzungberichte Academie Wissenschaft Wien*, **72**, 156-190.

Exner, S. (1888). Über optische Bewegungsempfindungen [On optical motion sensation]. *Biologisches Zentralblatt*, **8**, 437-448.

Farrell, J. E., & Shepard, R. N. (1981). Shape, orientation, and apparent rotational motion. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 477-486.

Francis, G., & Grossberg, S. (1996a). Cortical dynamics of boundary segmentation and reset: Persistence, after-images, and residual traces. *Perception*, **25**, 543-567.

Francis, G., & Grossberg, S. (1996b). Cortical dynamics of form and motion integration: Persistence, apparent motion and illusory contours. *Vision Research*, **36**, 149-173.

Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in spacetime. *Proceedings of the National Academy of Sciences*, **97**, 8186-8191.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Green, M. (1986a). Correspondence in apparent motion: Defining the heuristics. In *Proceedings of Vision Interface ´86*, pp. 337-342.

Green, M. (1986b). What determines correspondence strength in apparent motion? *Vision Research*, **26**, 599-607.

Green, M. (1989). Color correspondence in apparent motion. *Perception & Psychophysics*, **45**, 15-20.

Grossberg, S. (1994). 3-D vision and figure–ground segmentation by visual cortex. *Perception & Psychophysics*, **55**, 48-120.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, **92**, 173-211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, **38**, 141-171.

Grossberg, S., & Rudd, M. E. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review*, **99**, 79-121.

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics*, **43**, 241-277.

Hartmann, L. (1923). Neue Verschmelzungsprobleme [New fusion problems]. *Psychologische Forschung*, **3**, 319-396.

Hassenstein, B., & Reichardt, W. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus [A system-theoretical analysis of time-, sequence-, and sign-evaluation in the motion perception of the weevil chlorophanus]. *Zeitschrift für Naturforschung*, **11b**, 513-524.

He, Z. J., & Ooi, T. L. (1999). Perceptual organization of apparent motion in the Ternus display. *Perception*, **28**, 877-892.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1991). Focal visual attention produces motion sensation in lines. *Investigative Ophthalmology & Visual Science*, **22** (Suppl.), 144.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993a). Focal visual attention produces illusory temporal order and motion sensation. *Vision Research*, **33**, 1219-1240.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993b). Voluntary and stimulus-induced attention detected as motion sensation. *Perception*, **22**, 517-526.

Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge, MA: MIT Press.

Horowitz, T., & Treisman, A. (1994). Attention and apparent motion. *Spatial Vision*, **8**, 193-219.

Ivry, R., & Cohen, A. (1990). Dissociation of short- and long-range apparent motion in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 317-332.

Julesz, B. (1984a). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, **7**, 41-45.

Julesz, B. (1984b). Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Dynamic aspects of neocortical function* (pp. 585-611). New York: Neurosciences Research Foundation.

Kanizsa, G. (1951). Sulla polarizzazione del movimento gamma [On the polarization of gamma motion]. *Archivo di Psicologia, Neurologia e Psichiatria*, **3**, 224-267.

Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception*. New York: Praeger.

Kawahara, J., Yokosawa, K., Nishida, S., & Sato, T. (1996). Illusory line motion in visual search: Attentional facilitation or apparent motion? *Perception*, **25**, 901-920.

Kawamoto, M., Yoshino, A., Suzuki, A., & Ichinowatari, N. (1997). Human brain potentials observed using the line-motion method: The neurophysiological correlates of visual illusory motion perception. *Neuroscience Letters*, **231**, 49-52.

Kenkel, F. (1913). Untersuchungen über den Zusammenhang zwischen Erscheinungsgrösse und Erscheinungsbewegungen bei einigen sogenannten optischen Täuschungen [Investigations into the relationship between apparent size and apparent motion in several so-called optical illusions]. *Zeitschrift für Psychologie*, **67**, 358-449.

Klopfer, D. S. (1991). Apparent reversals of a rotating mask: A new demonstration of cognition in Perception. *Perception & Psychophysics*, **49**, 522-530.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Koffka, K. (1922). Perception: An introduction to the Gestalt theory. *Psychological Bulletin*, **19**, 531-585.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Köhler, W. (1959). Gestalt psychology today. *American Psychologist*, **14**, 727-734.

Kolers, P. A. (1972). *Aspects of motion perception*. New York: Pergamon Press.

Kolers, P. A., & Pomerantz, J. R. (1971). Figural change in apparent motion. *Journal of Experimental Psychology*, **87**, 99-108.

Kolers, P. A., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Research*, **16**, 329-335.

Korte, A. (1915). Kinematoscopische Untersuchungen [Kinematoscopic investigations]. *Zeitschrift für Psychologie*, **72**, 193-206.

Krumhansl, C. (1984). Independent processing of visual form and motion. *Perception*, **13**, 535-546.

Ledgeway, T., & Smith, A. T. (1994). Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Research*, **34**, 2727-2740.

Ledgeway, T., & Smith, A. T. (1995). The perceived speed of second-order motion and its dependence on stimulus contrast. *Vision Research*, **35**, 1421-1434.

Leeuwenberg, E. L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, **76**, 216-220.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, **7**, 3416-3468.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, **19**, 577-621.

Lorenceau, J., & Shiffrar, M. (1992). The role of terminators in motion integration across contours. *Vision Research*, **32**, 263-273.

Lu, Z.-L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, **35**, 2697-2722.

Mareschal, I., & Baker, C. L. (1998). Temporal and spatial response to second-order stimuli in cat A18. *Journal of Neurophysiology*, **80**, 2811-2823.

Mareschal, I., & Baker, C. L. (1999). Cortical processing of second-order motion. *Visual Neuroscience*, **16**, 527-540.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Marr, D., & Ullman, S. (1981). Direction selectivity and its use in early visual processing. *Proceedings of the Royal Society of London: Series B*, **211**, 151-180.

Mather, G., & West, S. (1993). Evidence for second-order motion detectors. *Vision Research*, **33**, 1109-1112.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gat-

tass, and C. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117-151). New York: Springer.

Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 130-138.

Navon, D. (1983). Preservation and change of hue, brightness, and form in apparent motion. *Bulletin of the Psychonomic Society*, **21**, 131-134.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, **8**, 2201-2211.

O'Keefe, L. P., & Movshon, J. A. (1998). Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Visual Neuroscience*, **15**, 305-317.

Olavarria, J. F., DeYoe, E. A., Knierim, J. J., Fox, J. M., & Van Essen, D. C. (1992). Neural responses to visual texture patterns in middle temporal area of the macaque monkey. *Journal of Neurophysiology*, **68,** 164-181.

Oram, M., & Perrett, D. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience*, **6**, 99-116.

Orlansky, J. (1940). The effects of similarity and difference in form on apparent motion. *Archives of Psychology*, **246**, 1-85.

Ousterhout, J. K. (1994). *TCL and the TK toolkit*. Reading, MA: Addison-Wesley.

Pantle, A. J., & Petersik, J. T. (1980). Effects of spatial parameters on the perceptual organization of a bistable motion display. *Perception & Psychophysics*, **27**, 307-312.

Pantle, A., & Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science*, **193**, 500-502.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, **82,** 2490-2502.

Perret, D., Harries, M., Mistlin, A. J., & Chitty, A. J. (1990). Three stages in the classification of body movements by visual neurons. In H. B. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding* (pp. 94-107). Cambridge: Cambridge University Press.

Petersik, J. T. (1995). A comparison of varieties of "second-order" motion. *Vision Research*, **35**, 507-517.

Petersik, J. T., & Pantle, A. J. (1979). Factors controlling the competing sensations produced by a bistable stroboscopic motion display. *Vision Research*, **19**, 143-154.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.

Ramachandran, V. S., Armel, C., Foster, C., & Stoddard, R. (1998). Object recognition can drive motion perception. *Nature*, **395**, 852-853.

Ramachandran, V. S., Ginsburg, A. P., & Anstis, S. M. (1983). Low spatial frequencies dominate apparent motion. *Perception*, **12**, 457-461.

Ramachandran, V. S., & Gregory, R. L. (1978). Does colour provide an input to human motion perception? *Nature*, **275**, 55-56.

Reichardt, W. (1961). Autocorrelation: A principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303-317). New York: Wiley.

Russell, B. (1965). Lectures on the philosophy of logical atomism. In R. C. Marsh (Ed.), *Bertrand Russell: Logic and knowledge: Essays 1901–1950* (pp. 178-281). London: Allen & Unwin (Original work published 1918)

Saito, H., Tanaka, K., Isono, H., Yasuda, M., & Mikami, A. (1989). Directionally selective response of cells in the middle temporal area (MT) of the macaque monkey to the movement of equiluminous opponent color stimuli. *Experimental Brain Research*, **75**, 1-14.

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in area MT: Effects on directional discrimination performance. *Journal of Neuroscience*, **12**, 2331-2355.

Schiller, P. (1991). Parallel pathways in the visual system: Their role in perception at isoluminance. *Neuropsychologia*, **29**, 433-441.

Seiffert, A. E., & Cavanagh, P. (1998). Position displacement, not velocity, is the cue to motion detections of second-order stimuli. *Vision Research*, **38**, 3569-3582.

Seiffert, A. E., & Cavanagh, P. (1999). Position-based motion perception for color and texture stimuli: Effects of contrast and speed. *Vision Research*, **39**, 4172-4185.

Shechter, S., Hochstein, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, **28**, 1013-1021.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, **91**, 417-447.

Shepard, R. N., & Zare, S. L. (1983). Path-guided apparent motion. *Science*, **220**, 632-634.

Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, **1**, 257-264.

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, **4**, 379-384.

Shiffrar, M., Li, X., & Lorenceau, J. (1995). Motion integration across differing image features. *Vision Research*, **35**, 2137-2146.

Shimojo, S., Silverman, G., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, **29**, 619-626.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, **384**, 460-463.

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudo type. *Journal of General Psychology*, **12**, 66-77.

Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology*, **38**, 168-172.

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *Journal of Neuroscience*, **18**, 3816-3830.

Steinman, R. M., Pizlo, Z., & Pizlo, F. J. (2000). Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research*, **40**, 2257-2264.

Stelmach, L. B., & Herdman, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 539-550.

Stelmach, L. B., Herdman, C. M., & McNeil, R. (1994). Attentional modulation of visual processes in motion perception. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 108-121.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 629-685). New York: Academic Press.

Stoner, G. R., & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, **5**, 129-149.

Ternus, J. (1938). The problem of phenomenal identity. In W. D. Ellis (Ed. and Trans.), *A source book of Gestalt psychology* (pp. 145-176). London: Kegan Paul, Trench, & Trubner. (Original work published 1926)

Titchener, E. B. (1908). *Lecture on the elementary psychology of feeling and attention.* New York: Macmillan.

Tse, P., & Cavanagh, P. (1995). Parsing occurs before line motion. *Investigative Ophthalmology & Visual Science*, **36**, 4.

Tse, P., & Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition*, **72**, 1-6.

Tse, P., Cavanagh, P., & Nakayama, K. (1996). The roles of attention in shape change apparent motion. *Investigative Ophthalmology & Visual Science*, **37**, 4.

Tse, P., Cavanagh, P., & Nakayama, K. (1998). The role of parsing in high-level motion processing. In T. Watanabe (Ed.), *High-level motion processing—Computational, neurobiological and psychophysical perspectives* (pp. 249-266). Cambridge, MA: MIT Press.

Tse, P., & Logothetis, N. K. (2001). [Transformational apparent motion occurs for equiluminant textures]. Unpublished raw data.

Ullman, S. (1978). Two dimensionality of the correspondence process in apparent motion. *Perception*, **7**, 683-693.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Ullman, S. (1988). Visual routines. In A. M. Collins & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 548-579). San Mateo, CA: Morgan Kaufmann.

Ullman, S. (1995). The visual analysis of shape and form. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 339-350). Cambridge, MA: MIT Press.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, & R. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A*, **1**, 451-473.

Victor, J. D., & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research*, **30**, 289-301.

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung [On the visually perceived direction of motion]. *Psychologische Forschung*, **20**, 325-380.

Wallach, H. (1976). *On perception*. New York: Quadrangle.

Wallis, G., & Bülthoff, H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, **98**, 4800-4804.

Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, **2**, 322-342.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, **20**, 158-177.

Werkhoven, P., Sperling, G., & Chubb, C. (1993). The dimensionality of texture-defined motion: A single channel theory. *Vision Research*, **33**, 463-485.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed. and Trans.), *A source book of Gestalt psychology* (pp. 71-88). London: Kegan Paul, Trench, & Trubner. (Original work published 1923)

Wertheimer, M. (1961). Experimental studies on the seeing of motion. In T. Shipley (Ed.), *Classics in psychology* (pp. 1032-1088). New York: Philosophical Library. (Original work published 1912)

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, **9**, 79-97.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus* (D. F. Pears & B. McGuinness, Trans.). London: Routledge (Original work published 1921)

Wuerger, S., Shapley, R., & Rubin, N. (1996). "On the visually perceived direction of motion" by Hans Wallach: 60 years later. *Perception*, **25**, 1317-1367.

Zanker, J. M. (1997). Is facilitation responsible for the "motion induction" effect? *Vision Research*, **37**, 1953-1959.

Zanker, J. M., & Huepgens, I. S. (1994). Interaction between primary and secondary mechanisms in human motion perception. *Vision Research*, **34**, 1255-1266.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, **236**, 549-573.

Zhou, Y.-X., & Baker, C. L. (1993). A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science*, **261**, 98-101.

Zhou, Y.-X., & Baker, C. L. (1994). Envelope-responsive neurons in Area 17 and 18 of cat. *Journal of Neurophysiology*, **72**, 2134-2150.

Zhou, Y.-X., & Baker, C. L. (1996). Spatial properties of envelope responses in Area 17 and 18 of the cat. *Journal of Neurophysiology*, **75**, 1038-1050.

APPENDIX
Summary of Experimental Questions and Results for All Experiments

Experiment	*n*	Question Addressed	Answer
1	8	Can matching in translational apparent motion be based solely on color?	No.
2	2	What is the temporal boundary between continual and halting motion in the TAM sequences used in Experiment 3?	TAM appears continual when the biasing image <~216 msec.
3		What is the duration that a form must be present to bias TAM in the vertical or horizontal direction using a modified quartets stimulus?	
3A	8	Red and green on a black background (not equiluminant).	75% threshold; 126 msec
3B	8	Red and green on a gray background (equiluminant).	75% threshold: 106 msec
3C	7	Black outline squares and rectangles on a white background.	75% threshold: 126 msec
3D	7	White outline squares and rectangles on a black background.	75% threshold: 88 msec
4		Is TAM consistent with an existing 3-D interpretation over an ambiguous stimulus?	
4A	3	An ambiguous Necker-like stimulus grows and stops.	Yes. >95% TAM consistent
4B	3	Same as Experiment 4A, but stimulus grows and contracts.	with existing 3-D shape percept.
5	10	How long must an unambiguous 3-D form be present in order to bias 3-D TAM?	80–140 msec
6	10	Were results of Experiment 5 due to afterimages? Control for Experiment 5.	No. >80% TAM in biased direction. Results of Experiment 5 not due to afterimages.
7	6	Were results of Experiment 5 due to the solid stimulus biasing successive Necker stimuli rather than to TAM? Control for Experiment 5.	No. The results of Experiment 5were due to TAM and not form biasing.

Note—Experiments 3 and 5 were the most important experiments.

(Manuscript received October 23, 2000;
revision accepted for publication April 24, 2001.)