

Attention-biased multi-stable surface perception in three-dimensional structure-from-motion

Karel Hol

Helmholtz Institute, Utrecht University
Utrecht, The Netherlands

Ansgar Koene

Espace et Action, Institut National de la Santé et de la
Recherche Médicale, Bron, France

Raymond van Ee

Helmholtz Institute, Utrecht University
Utrecht, The Netherlands



Retinal velocity distributions can lead to a percept of three-dimensional (3D) structure (structure-from-motion [SFM]). SFM stimuli are intrinsically ambiguous with regard to depth ordering. A classic example is the orthographic projection of a revolving transparent cylinder, which can be perceived as a 3D cylinder that rotates clockwise and counterclockwise alternately. Prevailing models attribute such bistable percepts to inhibitory connections between neurons that are tuned to opposite motion directions at equal binocular disparities. Cylinder stimuli can yield not only two but as many as four different percepts. Besides the well-documented clockwise and counterclockwise spinning transparent cylinders, observers can also perceive two transparent half-cylinders, either convex or concave, one in front of the other. Observers are able to bias the time during which a percept is present by attending to one or the other percept. We examined this phenomenon quantitatively and found that in standard SFM stimuli, the percept of two convex transparent half-cylinders can occur just as often as the percept of (counter-) clockwise spinning cylinders. So far, however, all interpretations of experimental (neurophysiological) data and all proposed mechanisms for SFM perception have focused solely on the two classical cylinder percepts. Prevailing models cannot explain the existence of the other two percepts. We suggest an alternative model to explain attention-biased multi-stable perception.

Keywords: attention, structure-from-motion, shape, depth

1. Introduction

It has long been known that time-varying two-dimensional (2D) images can evoke a strong perception of structure and motion in depth, even in the absence of other depth cues (Miles, 1931; Wallach & O'Connell, 1953). The ability to see the three-dimensional (3D) structure of objects from motion cues alone (structure-from-motion [SFM]) has been extensively studied by using computer-generated moving random-dot patterns (RDPs) (for a review, see Andersen & Bradley, 1998). In the laboratory, a frequently used SFM display is one that represents the orthographic projection of a revolving transparent cylinder by superimposing two sets of randomly positioned dots moving in opposite directions (see Figure 1). Each RDP moves with a sinusoidal speed profile (i.e., dots in the middle of the display move at a high speed that drops to zero at the display's edges). When viewing SFM stimuli, the depth information can only be recovered on the basis of motion measurements. In the resulting percepts, dots moving rightwards might initially appear to be in front and the leftward-moving dots in the back; this translates into a transparent cylinder that appears to rotate counterclockwise (from the top). After a matter of seconds, the perceived rotation direction

reverses. This kind of SFM perception is said to be "bistable" (much like a Necker cube, whose perceived 3D structure reverses). The stimulus can be rendered unambiguous by the addition of binocular disparities that define the dots' depth order (e.g., Braunstein, Andersen, Rouse, & Tittle, 1986).

There is a large body of literature on how the brain uses motion information to perceive SFM (e.g., Husain, Treue, & Andersen, 1989; Doshier, Landy, & Sperling, 1989; Treue, Husain, & Andersen, 1991; Treue, Andersen, Ando, & Hildreth, 1995). From electrophysiological studies, we know that cortical middle temporal area (MT) in the macaque monkey is specialized for processing visual motion information (Maunsell & Van Essen, 1983a). MT neurons are also selective to binocular disparity (Maunsell & Van Essen, 1983b; Qian & Andersen, 1994; Bradley, Qian, & Andersen, 1995; DeAngelis & Newsome, 1999). Furthermore, electrical microstimulation in MT influences the perceptual responses of monkeys in both motion (Newsome & Paré, 1988; Newsome, Britten, & Movshon, 1989) and stereo tasks (DeAngelis, Cumming, & Newsome, 1998). Bradley et al. (1995) reported that MT neurons are selective for transparent surface movements at different disparity planes. More recently, Bradley, Chang, and Andersen

(1998) and Dodd, Krug, Cumming, and Parker (2001) reported that the activity of MT neurons correlated with monkeys' perceptual responses in a depth order task. In these studies, monkeys viewed transparent rotating cylinders defined by SFM. The monkeys were trained to indicate the motion direction of the perceived cylinder's front surface. When disparity cues were present in the stimulus, the direction indicated was consistent with the stimulus' attributes (motion direction combined with disparity information). When disparity was absent (i.e., the depth ordering was ambiguous), MT responses were correlated with the reported rotation direction. For example, when the monkey indicated that the near surface was moving rightwards, MT cells preferring rightward motion and near disparities and those preferring leftward motion and far disparities were more active than those neurons having opposite motion tuning at a given disparity. If the monkey indicated that it perceived that the front surface was starting to move in the opposite direction, then the near-left and far-right cells would become active and the others would become less active. These results reflect a strong correlation between the neuronal responses and the animals' response about the motion direction of the perceived front surface.

1.1 Two Alternative Percepts

Existing models and, more importantly, the (neurophysiological) interpretations of the experimental data rely largely on the assumption that SFM stimuli are essentially bistable (i.e., only one of two different percepts is present at a time). Notice, however, that when an RDP is presented with only one motion direction and a sinusoidal speed profile, dots appear to be on the cylinder's convex surface facing the observer (Nawrot & Blake, 1991a). As the dots move, they seem to come closer to the observer, only to disappear as they pass behind what looks like the cylinder's border. Note that a second geometrically plausible percept is a transparent half-cylinder with dots visible only on its back surface. By the same token, when two dot patterns are superimposed, there are four possible percepts. In addition to the two well-documented clockwise and counterclockwise spinning transparent cylinders, there are two other possible percepts. The stimulus can also look either like two convex ("two-fronts" percept; Figure 1g) or like two concave ("two-backs" percept; Figure 1h) transparent half-cylinders with surface motion in opposite directions.¹ Thus the SFM display is more variable than in traditional models (Koene, Hol, & van Ee, 2002). In addition, different attentional states could bias the probability of seeing a particular percept. Although quite a few researchers seem to be aware of the attention-biased four-percept phenomenon, it does not seem to have been formally reported. Moreover, no quantitative studies exist in the literature on the

mentioned four-percept phenomenon, and the alternative percepts cannot be predicted or explained by prevailing models.

1.2 Aim of the Study

We intended to gain further insights into the process controlling the percept changes related to this SFM stimulus and the role that attention plays in this process. Our expectation that attention might play an important role in controlling the percept was in part motivated by the results of previous neurophysiological studies. Single-unit and neuroimaging experiments (Brefczynski & DeYoe, 1999; Gandhi, Heeger, & Boynton, 1999; Somers, Dale, Seiffert, & Tootell, 1999; Treue & Maunsell, 1996) have shown that directed spatial attention and featural attention to motion (Beauchamp, Cox, & DeYoe, 1997; Chawla, Rees, & Friston, 1999; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Treue & Martinez Trujillo, 1999) can increase responses in both human MT+ and monkey MT.

Instructions to observers and their experience play an important role in the perception of SFM stimuli. By means of both pilot experiments and demonstrations in front of audiences, we found that practically all observers have very little difficulty in perceiving both the cylinder and the two-fronts percepts after these percepts have been explained to them. However, we found that only observers who are highly experienced with this kind of SFM stimulus are able to experience the two-backs percept. We therefore concentrated on further analyzing the two-fronts percept and the cylinder percepts. In our experiments, we cued observers to attend either to the percept of a transparent cylinder (Figure 1f) or to the percept of two convex transparent half-cylinders ("two-fronts"; Figure 1g), and to report the presence of the percept. It should be stated, however, that cueing to the two-fronts percept is not a prerequisite for experiencing the two-fronts percept. During demonstrations, we encountered a considerable number of observers who perceived both the cylinders and the two-fronts percepts spontaneously even before they had been informed about the possible percepts.

As we will show, our results indicate that the two-fronts percept was just as quickly present as the cylinder percept, and both percepts were equally stable. To gain insight into the mechanisms underlying these percepts, we compared our data with predictions based on the prevailing SFM model of Andersen and Bradley (1998). Because this seminal model focuses on the two classical cylinder percepts, it cannot adequately account for our experimental results. We therefore provide an alternative model to explain the attention-biased multi-stability in 3D SFM perception.

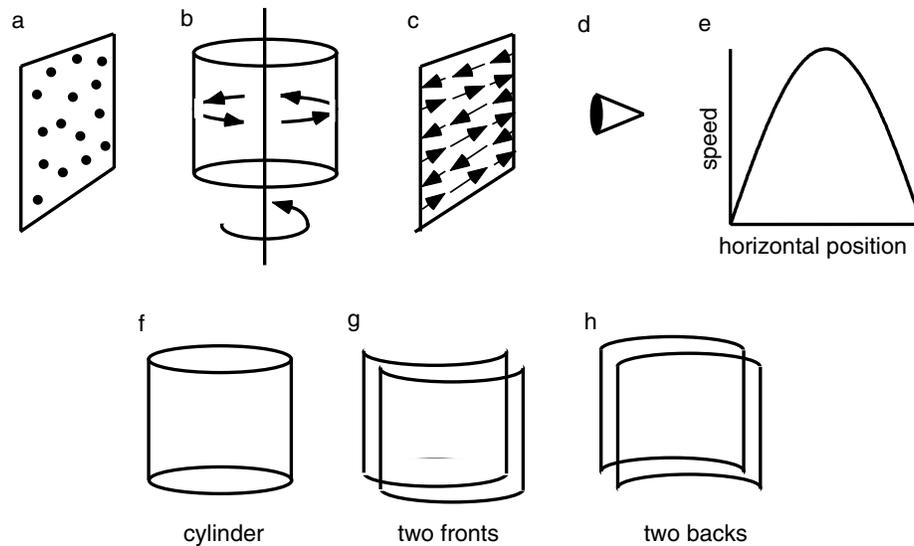


Figure 1. Diagrammatic representation of a typical structure-from-motion (SFM) stimulus that generates the percept of cylindrical shapes. To construct the stimulus, dots are randomly plotted on a 2D square (a), and projected onto a transparent surface of a rotating cylinder (b). c shows the resulting stimulus as presented to the observer (d) when the individual dot speed is varied as a half-sinusoid across the display (e) with the highest speeds in the center. Different attentional states render the SFM display more unstable than is traditionally assumed; it allows not only two but four different percepts: a clockwise or counterclockwise rotating transparent cylinder (f), two convex (g) or two concave (h) transparent half-cylinders spinning in opposite directions. Demonstrations of our stimuli can be seen at <http://www.phys.uu.nl/~vanee/tube.html>.

2. Methods

2.1 Stimuli

We simulated a parallel projection (without perspective cues; Treue et al., 1991) of a transparent rotating cylinder covered with dots onto a flat plane. Figure 1a illustrates the stimulus construction. Moving patterns of yellow dots on a black background were generated on an Apple Macintosh G4 computer and displayed on a color CRT monitor (LaCie) in a dimly lit room. Using a chin rest and a head support, observers were seated 114 cm away from the monitor. Monitor resolution was 59.7 pixels per degree of visual angle and its refresh rate was 74.6 Hz. The number of dots presented in each frame was 138 (69 for each surface viz. motion direction), resulting in a dot density of 5.9 dots deg^2 , where individual dots subtended 3 arcmin. The lifetime of each dot was infinite. In order to simulate rotation, the dots moved with a sinusoidal speed profile (i.e., stimulus features in the middle of the display moved at a high speed that dropped to zero at the display's edges) (Figure 1e). Dots moving in a certain direction (leftward or rightward) were wrapped around when they reached the edge of the stimulus. The stimulus diameter was 3.5° and its height 7.0° . Dots moved by 1° every two frames (27 ms) (i.e., the maximum speed of the dots was 37.5°s^{-1}). The sequencing of animation frames was synchronized to the monitor's refresh rate and was completed in a single interrupt, ahead of the beam trace. This careful stimulus

presentation ensured uniform temporal parameters for the stimulus. Special care was taken to prevent occlusion from dots with opposing motion directions by making a corrugated display (i.e., dots moving in opposite directions were confined to odd and even numbered vertical stacks). The gap between stacks could be varied between experimental runs though it was typically kept constant at 0.1° .

2.2 Design

Before any data were collected, all observers participated in a series of 20 trials. We first showed a unidirectional flow of coherent motion (opaque front surface) followed by the bidirectional flow. Except for the authors, observers were unfamiliar with the possible percepts these stimuli could evoke. The possible percepts were explained to the subjects by hand movements indicating the shapes the stimulus could take. When cued, all observers, naïve and experienced, were able to perceive both the cylinder and the two-fronts percepts in this first training run.

Each trial started with a key press. Observers initiated trials at their own pace. No feedback was given. Sessions consisted of 20 trials, lasting 20 s each. Observers viewed the display with both eyes. Attention played a paramount role in our experiments. In most of the experimental manipulations, observers were cued before the start of a trial to attend either to the cylinder percept or to the two-fronts percept in a randomly interleaved fashion by

displaying the string “CYL” or “2F,” respectively. Then the stimulus was shown and the observer was required to press and hold down a response key for as long as the cued percept was present. Each stimulus condition was presented five times during each session.

In a blocked design, we tested the following four conditions.

2.2.1 Spacing between stacks

We hypothesized that the two-fronts percept could be an artifact due to occlusion effects. When front and back surfaces of a 3D object are projected onto a flat screen, these front and back surfaces should not be projected on top of each other to prevent occlusions between dots. A vertical spacing between horizontal rows of dots preserves the observer’s perspective (Figure 2a). We varied the spacing between rows of dots from 0° to 0.1° to 0.2° in different experimental sessions. The 0.2° condition corresponds to the realistic situation in which, for the used viewing distance, a row of dots in the front surface would never occlude the row of dots in the back surface. Because this manipulation did not yield a significant effect, we used a spacing of 0.1° in the subsequent experiments.

2.2.2 Attention to a motion direction

In studies concerned with the neuronal correlates of 3D SFM perception, subjects are required to report the motion direction of the perceived front surface. To make a comparison with these studies, we examined the role of attending to motion direction. Unlike depth order, motion direction is unambiguous. Before a particular session, observers were instructed to attend either to the cylinder or to the two-fronts percept. In a session, they were cued before the start of a trial (i) to disregard the motion direction, or (ii) to attend to the leftward, or (iii) to attend to the rightward motion direction (by displaying the symbol ‘<’, ‘>’ or ‘>’) in a randomly interleaved fashion. Note that both of the latter conditions required observers to attend simultaneously to a shape (cylinder or two-fronts) and to a motion direction.

2.2.3 Fixation

To examine the impact of eye movements on our results, we tested observers in three different conditions: (i) free viewing, (ii) fixation of a non-occluded white square at the display’s center, or (iii) fixation of a white cross presented 2.93° to the left of the stimulus center.

2.2.4 Relative disparity

In this final condition, we added relative horizontal disparity to the stimulus by means of a conventional anaglyph (red/green) setup. Horizontal disparity was used to separate the two surfaces in depth. The disparity information conveys the percept of a transparent cylinder whose depth order is unambiguous (e.g., Braunstein et al., 1986). Disparity was added so that each moving surface

received equal but opposite disparity. The disparity of each individual dot was scaled according to its distance from the midline, in a manner similar to the speed scaling described above. Thus, the maximum disparity occurred at the stimulus’ center and decreased toward the edges. A positive disparity specified near depth (crossed disparity) to rightward moving dots and far depth (uncrossed disparity) to leftward moving dots (counterclockwise rotation as viewed from above). A negative disparity specified near depth to leftward moving dots and far depth to rightward moving dots (clockwise rotation). The amplitude of the disparities used corresponded to a cylinder of zero diameter (0), half diameter (± 0.5), or full diameter (± 1). For large disparities, the direction of rotation is defined unambiguously (e.g., Braunstein et al., 1986). For human and monkey observers, large disparities produce nearly perfect performance (Dodd et al., 2001). As disparities are reduced, discrimination between clockwise and counterclockwise rotation becomes increasingly difficult, and with increasing frequency observers perceive the rotation opposite to that defined by the sign of disparity. We also added disparity to the stimulus such that it specified two convex transparent half cylinders (two-fronts). In this case, a positive disparity specified near depth to both leftward and rightward moving dots. In each set of trials (36 per experimental session), the magnitude (0, ± 0.5 , and ± 1) and type of disparity information (specifying a cylinder or to a two-fronts percept) were randomly interleaved.

2.3 Data Analysis

For all conditions, individual observer data were averaged across five trials per condition. Individual observers’ mean values were averaged across observers (usually four, see Section 2.4). We analyzed the length of time during which a certain percept was present, together with the observers’ reaction times. Reaction times were quantified as the time from stimulus onset until the moment when observers reported perceiving the cued percept for the first time in the trial. These times were used to evaluate the build up of percepts. Statistical significance was assessed by two-tailed paired Student *t* test for means.

2.4 Subjects

Eleven observers participated. Seven of them were unfamiliar with SFM stimuli and were naïve concerning the purpose of the experiment. Four observers were experienced with SFM stimuli (two of them, RE and KH, were authors). Two of the observers were asked to participate specifically because they had reported seeing the two-backs percept in pilot experiments. These two observers (AD and SP) were highly experienced in viewing SFM stimuli. All 11 observers participated in the experiments reported in Section 3.1 and Section 3.2. In

all other conditions, a fixed, four-member subgroup of these 11 observers participated; two were authors and the remaining two were naïve. All observers had normal or corrected-to-normal vision and good stereo acuities. The observers who participated in the disparity conditions completed a metrical stereo test (van Ee & Richards, 2002). These observers were able to distinguish disparities that had different signs and magnitudes within the fusional range.

3. Results

When the stimulus was a single dot pattern with a sinusoidal speed profile observers perceived a convex cylindrical surface, in agreement with previous results (Nawrot & Blake, 1991a). When both leftward and rightward moving dot patterns were presented simultaneously on top of each other in a corrugated way, the stimulus evoked four different percepts: the well-documented counterclockwise or clockwise transparent cylinders, and two convex (or to a lesser extent concave) transparent half cylinders with surface motion in opposite directions, one of the surfaces appearing to be nearer the observer than the other. Even though the observers perceived that both rightward and leftward moving dot patterns had the same curvature, they perceived the surfaces to be close but at different depths. As noted above, the two-backs percept was seen only by highly experienced observers.² We therefore analyzed the two-fronts percept and the cylinder percepts. We tested five different conditions.

3.1. Spacing Between Stacks

We varied the spacing between horizontal rows of dots (stacks) from 0° to 0.1° to 0.2° in different experimental sessions. The total length of time (averaged across 11 observers) during which a percept was reported to be perceived is shown in Figure 2b as a function of the spacing between vertical stacks of dots. Figure 2c shows observers' reaction times as a function of the spacing between vertical stacks of dots. Both the total period during which a stable percept was present and the reaction times remained constant over the different spacing values used for both cylinder and two-fronts percepts. The two-fronts percept is therefore not an artifact due to occlusions between rows of dots. Notice, however, that the cylinder was perceived earlier and for a slightly longer time than the two-fronts, irrespective of the spacing between stacks. Furthermore, the two-fronts percept (as well as the cylinder) was present for more than half the presentation time. Taking this result into account, we proceeded using the 0.1° inter-stack interval as the standard spacing between rows of dots.

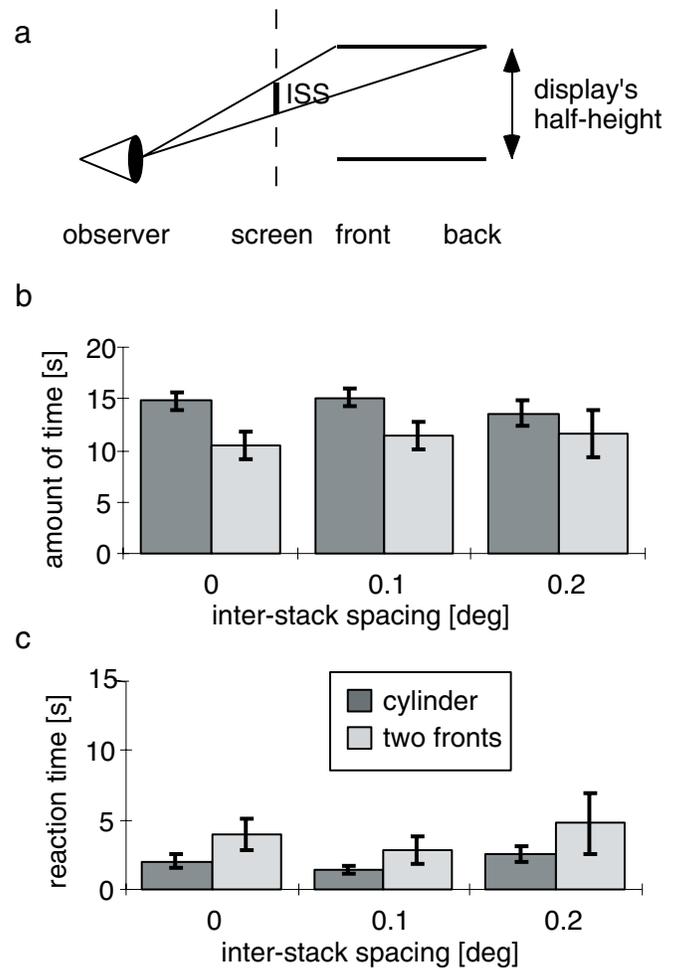


Figure 2. When front and back surfaces of a 3D object are projected onto a flat screen, a vertical spacing should be present between dots in order to preserve the observer's perspective (a). The horizontal lines represent the stimulus diameter (from front to back surfaces). The vertical line (ISS) represents the distance between two dots (pixels) of the projected front and back surfaces, at the same height. b shows the summed time during which a percept (cylinder or two fronts) was present during the 20-s presentation period, as a function of the spacing between vertical stacks. Before the actual stimulus was shown, subjects were cued in a random interleaved fashion to attend either to the (clockwise or counterclockwise) cylinder (dark bars) or to the two-fronts (light bars). The results indicate that the two percepts were present for a constant amount of time irrespective of the spacing values used, indicating that the two-fronts percept is not due to occlusions between dots. c shows the reaction time taken to perceive the cued percept, as a function of the spacing between vertical stacks. The reaction times remained constant across spacing values used for both cylinder and two-fronts percepts. The mean of the five trials per condition for 11 individual observers was averaged. Error bars show the $\pm 95\%$ confidence intervals.

3.2 Attention to a Motion Direction

Figure 3a and Figure 3b show the total amount of time during which a particular shape was perceived during the trial and the reaction times as a function of the attentional condition (attend to none, leftward or rightward motion), respectively. This manipulation did not result in different percepts or in different reaction times for either the cylinder or the two-fronts percepts. Thus, irrespective of whether observers were attending to a motion direction or not, they could perceive cylinders and two-fronts for more than half of the presentation time.

3.3 Fixation

Given that the two-fronts percept was not caused by occlusions or by difficulties in devoting attentional resources to the motion direction (an unambiguous

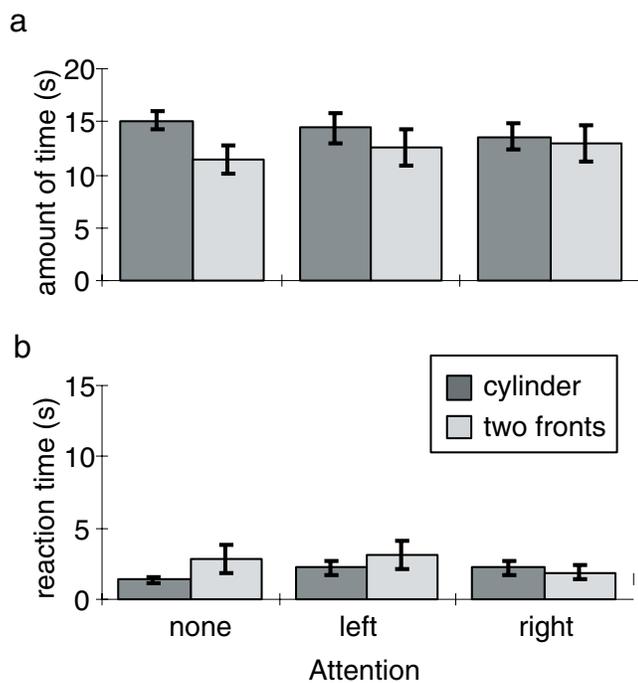


Figure 3. In blocked sessions observers were cued verbally to attend to the (clockwise or counterclockwise) cylinder or to the two-fronts. Within an experimental session they were cued in a randomly interleaved fashion to disregard the motion direction (none) or to leftward or rightward motion. Left: attending simultaneously to leftward motion together with attending either to a cylinder or two-fronts. Right: attending simultaneously to rightward motion together with attending either to a cylinder or two-fronts. a shows the summed time during which a percept (cylinder or two fronts) was present. b shows the reaction times for the perception of a cylinder (dark bars) or two-fronts (white bars), as a function of attention to a given direction of motion. The data indicate that attention to motion direction does not affect the total duration of either percept nor does it influence the reaction times.

feature of the display), we examined whether eye movements could have created it. We studied three different conditions in blocked-design: (i) free viewing, (ii) fixating a non-occluded white square in the center of the display, and (iii) fixating a non-occluded white cross presented eccentrically to the left of the patterns' center.

Figure 4a shows the total time during which a shape (cylinder or two-fronts) was perceived per 20 s trial as a function of the fixation condition. Figure 4b shows reaction times as function of the fixation condition. Irrespective of the fixation condition, the cylinder was perceived for a similar length of time, and the reaction times were also similar. The length of time during which the two-fronts were perceived was, however, affected by

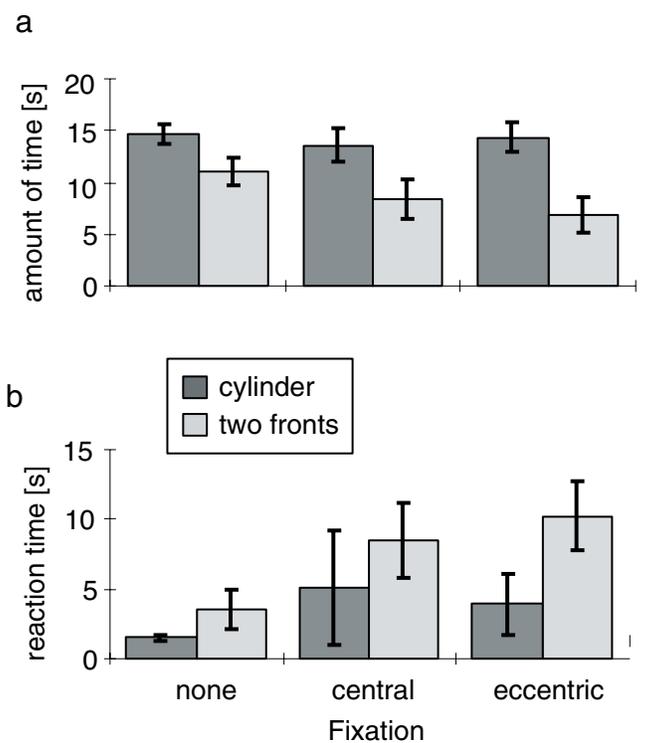


Figure 4. The role of fixation location. As in Figure 2, observers were cued to attend to a cylinder or to two-fronts in a randomly interleaved fashion. In a block of trials, fixation was either free (none), central, or eccentric (2.93° to the left of the stimulus). a shows the total time during which a percept was present, and b shows the reaction times for perceiving a cylinder or two-fronts, as a function of fixation position. The figure shows that the periods of time during which the cylinder was present and the reaction times were constant across the different fixation conditions. In contrast, the duration of the two-fronts percept was shortest for the eccentric fixation condition, and the reaction times increased when fixation was eccentric. In this condition, the time during which the cylinder percept was present was almost twice as long as the time during which the two-fronts percept was present. Differences in the reaction times for reporting the presence of the two percepts were also largest in the latter condition.

fixation, the time being shortest for the eccentric fixation condition. Reaction times increased as the fixation location varied from none to central to eccentric. For the eccentric fixation condition, the time during which the cylinder was perceived was almost twice as long as the time during which the two-fronts were perceived. Reaction times behaved in the opposite way (i.e., reaction times for reporting the two-fronts were almost twice as long as the reaction times for reporting the percept of a cylinder).

3.4 Relative Horizontal Disparity

Horizontal disparity was used to separate the two surfaces in depth, thereby unambiguously specifying the direction of rotation. Trials in which the disparity information specified the percept of either a cylinder or two-fronts were randomly interleaved. We will first focus on the disparities that specified a cylinder.

3.4.1 Relative disparity specifying a cylinder

Horizontal disparity was added so that each moving surface received equal but opposite disparity. Hence, the center of the 3D transparent cylinder corresponded to the monitor plane. The disparity information specified the percept of a cylinder. Five different relative disparities were used: zero (as in the experiments so far), ± 0.5 , and ± 1 , with 1 being the situation in which the maximum depth between front and back surfaces equals the stimulus' width, making the predominant percept a counterclockwise spinning cylinder. Figure 5 shows the time during which a shape was perceived as a function of relative disparity. As disparity increased from 0 to ± 1 , the

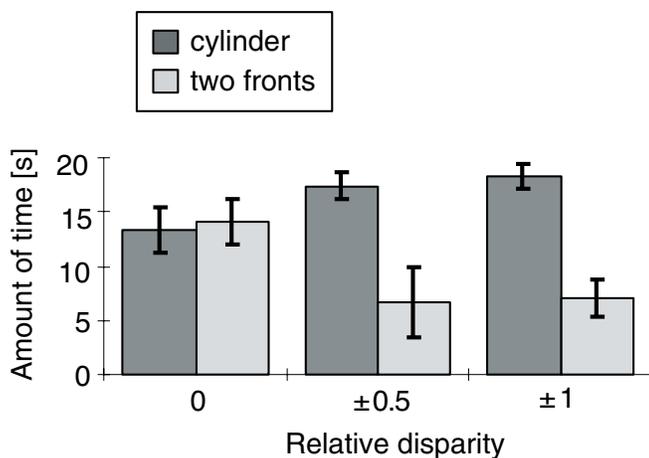


Figure 5. Variation of the relative horizontal disparity specifying the cylinder percept. The amplitude of the disparity signal corresponded to a cylinder of zero-diameter (0), half diameter (± 0.5), or full diameter (± 1). Observers were cued to attend to a cylinder or to two-fronts. As disparity increased, the total duration of the two-fronts percept decreased, though even with a disparity of ± 1 observers still reported seeing the two-fronts for a considerable period of time.

time during which the two-fronts were perceived decreased. Notice, however, that even with a relatively large disparity, observers still reported seeing the two-fronts for more than a quarter of the presentation times.

3.4.2 Relative disparity specifying two-fronts

Next we focus on the disparity information specifying the percept of two-fronts; these trials were randomly interleaved with those having disparity specifying the percept of a cylinder. Three different relative disparities (the positive ones used in the aforementioned condition) were used in this condition: zero, $+0.5$, and $+1$. Figure 6 shows the time during which a shape was perceived as a function of relative disparity specifying the two-fronts percept. Independent of disparity, the two percepts (cylinder and two-fronts) did not differ significantly with regard to the time during which they were perceived. Notice that even with a relatively large disparity specifying two-fronts, observers still reported seeing the cylinder for more than half of the presentation time.

3.4.3 Comparison between the two disparities used

A comparison between the data shown in Figure 5 (disparity specifying a cylinder) and Figure 6 (disparity specifying two-fronts) shows that across disparity magnitudes, the cued shape was perceived for a longer time than the other shape, when the disparity information specified that shape. The cylinder percept was, however, not as strongly affected by the disparity as was the two-fronts percept.

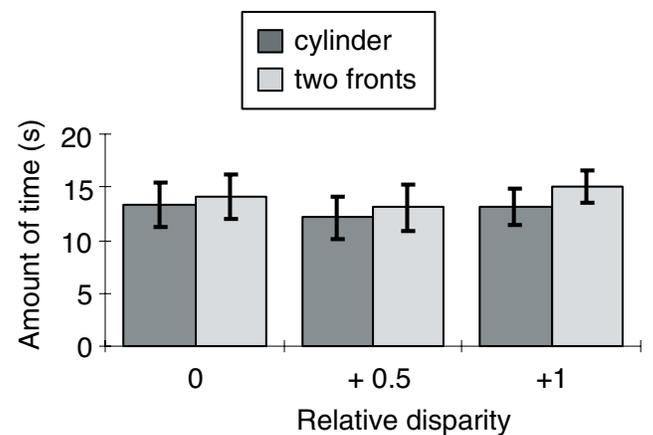


Figure 6. Results when the relative horizontal disparity specified the two-fronts percept. As in Figure 5, the amplitude of the disparity signal corresponded to zero-diameter (0), half diameter ($+0.5$), or full diameter ($+1$). Observers were cued to attend to a cylinder percept or to the two-fronts percept. Independent of disparity, the two percepts were present for the same length of time. Even with a disparity of $+1$, observers still perceived the cylinder for more than half of the presentation time.

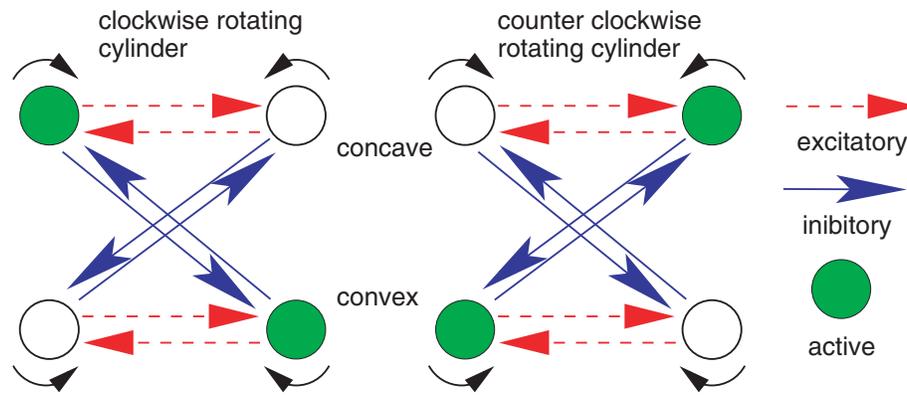


Figure 7. Schematic representation of the two stable states in the circuit of direction and disparity-dependent interaction in MT that were proposed by Andersen and Bradley (1998). The circles represent MT neurons coding for convex or concave depth curvature (indicated by the black arrows). The filled circles are active neurons; the empty ones are inactive. The arrows between the neurons indicate excitatory or inhibitory connections.

4. Model

To place our results in the context of current knowledge of how SFM percepts arise, we compared our data to the predictions of the prevailing SFM model by Andersen and Bradley (1998). This model was designed to explain the spontaneous changes in perceived depth curvature caused by SFM stimuli, such as those used in our experiment. In the Andersen and Bradley (1998) model, the percept of a clockwise or counterclockwise spinning cylinder is related to the state of a bi-stable network of MT neurons (Figure 7).

In a noise-free system, the ambiguous SFM stimulus would activate all the MT neurons in this network equally. Due to signal noise, however, some neurons are activated slightly more than others. The excitatory and inhibitory connections between the neurons amplify this slight difference and cause the network to move into one of the two stable states (Figure 7). The modeled system then remains in this state until neural fatigue causes the network to shift to the other stable state (Andersen &

Bradley, 1998).

The presence of only two possible stable percepts is incompatible with our experimental results, which revealed four stable percepts. In order to explain our experimental findings, we therefore propose an alternative model for the perception of depth curvature from SFM. There are two key differences between our model and the one of Andersen and Bradley; first, we assume that the depth curvatures evoked by rightward and leftward moving dots arise independently. Second, the stability of the percept is not based on a bi-stable network but is due to temporal integration (i.e., low-pass filtering). A model similar to the one we propose was developed earlier by Taylor and Aldridge (1974) to model the reversal between convex and concave percepts of the dents in an otherwise flat surface.

4.1 An alternative model

The structure of our model is shown in Figure 8. First, motion detectors are activated by the rightward and leftward moving dots. The selectivity of the motion detectors results in a “common fate” separation of the

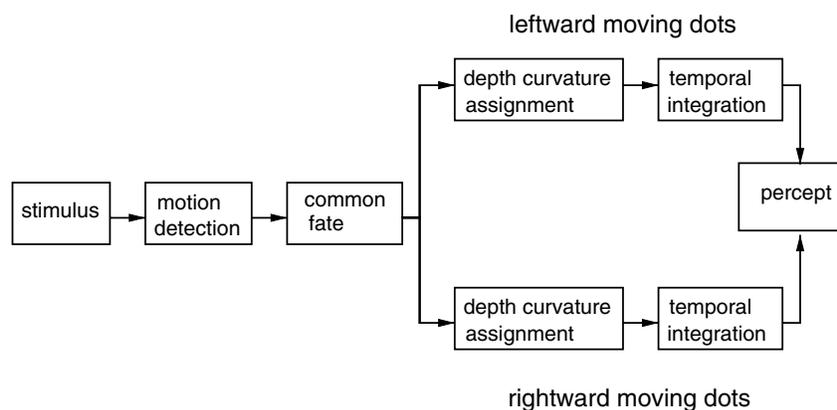


Figure 8. The functional stages of our model. From left to right, the five computational processing stages that lead to the percept of the SFM stimulus are shown.

dots: Dots moving in one direction are processed by one population of MT neurons while dots moving in another direction are processed by a different population of MT neurons (DeYoe & Van Essen, 1988; Poggio, Gonzalez, & Krause, 1988; Livingstone & Hubel, 1987; Zeki, 1974). In the next stage, the sinusoidal speed profiles lead to a depth curvature assignment (i.e., convex or concave half cylinders) (Fernandez, Watson, & Qian, 2002). Because the depth curvature assignment occurs for the leftward and rightward moving dots independently, all four permutations of the pairs (left, right) (convex, concave) are equally possible (the effect of differences in the prior probabilities of convex and concave depth curvature assignment will be addressed in the “Discussion”). For most natural stimuli, depth cues, such as disparity, perspective, and so on, determine the polarity of the depth curvature (convex or concave). Because none of these cues is present in a standard SFM stimulus, the depth curvature polarity is ambiguous. Under these conditions, the assigned depth curvature will be the result of signal noise and therefore change stochastically (Merk & Schnakenberg, 2002). Figure 9a shows the temporal dynamics of this initial depth curvature assignment stage. If we assume that the highest frequency of percept changes gives us the rate at which the depth curvature assignment is updated, then the addition of a temporal integration over an appropriate time window is needed so the model can produce the longest percept durations (which were found to go up to 5 s). The output of the temporal integration stage simply corresponds to the curvature that, during the period inside the integration window, has been assigned more often to the leftward/rightward moving dots. For stimuli with unambiguous depth cues, the temporal integration

ensures shape constancy by filtering out signal noise. Figure 9b illustrates the temporal integration by showing the difference in the number of times that convex or concave curvatures have been assigned ($\# \text{ convex} - \# \text{ concave}$) as function of time. Figure 9c shows the resulting output, and Figure 9d shows the temporal dynamics of the corresponding percepts.

5. Discussion

In summary, we have reported quantitative experimental results regarding two phenomena: (1) SFM displays resulting from the parallel projection of a transparent cylinder allow for four rather than the two traditionally reported percepts. (2) Observers were able to attentively bias the average period of time during which the different percepts were present.

Observers were able to perceive not only the well-known clockwise or counterclockwise rotating transparent cylinders (Figure 1f), but also two convex (two-fronts, Figure 1g) and to a lesser extent two concave (two-backs, Figure 1h) transparent half cylinders with surface motion in opposite directions. One of these transparent half cylinders appeared to be nearer the observer than the other one, even though observers perceived that both rightward and leftward moving patterns had the same curvature. The two-backs percept was perceived only by a few experienced SFM observers. The two-fronts percept was as stable and readily available as the percept of a cylinder, and the strength of this percept was only diminished when observers were required to fixate.

Although we demonstrated that attending to a direction of motion did not cause any changes either in the percept's presence over time or in the depth order, we

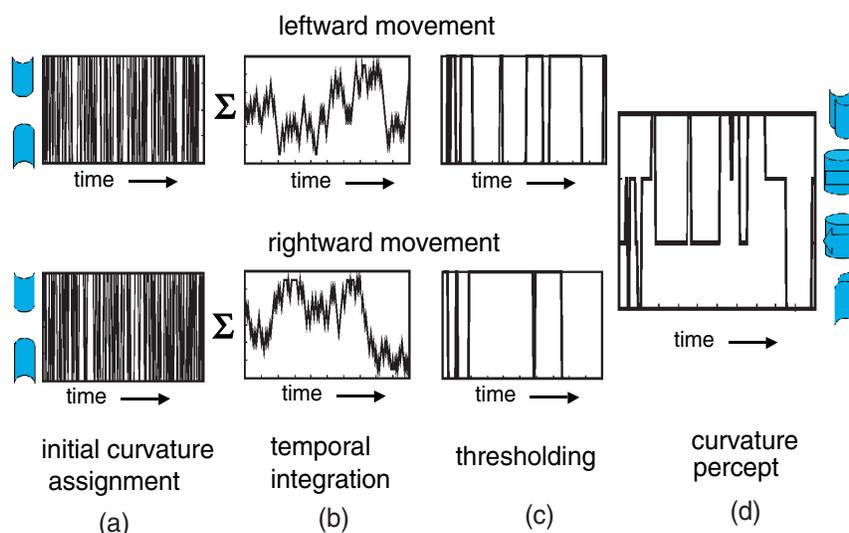


Figure 9. Temporal dynamics of the assigned depth curvature. The graphs indicate the curvature assignment (y-axis) as a function of time (x-axis). Σ indicates a summation of the inputs over time. From left to right, we show how the process of temporal integration (b) reduces the rapid stochastic curvature changes in the initial curvature assignment (a) to a more stable state with, on average, fewer assignment changes (c). Finally, d shows the temporal changes in the perceived shape of the SFM stimulus.

showed that attention to shape could influence the percept of 3D SFM displays. Existing models for SFM perception (Andersen & Bradley, 1998; Nawrot & Blake, 1991b) do not take attentional effects into account. We examined whether both the time it takes for the observer to become aware of the percept (quantified by the reaction times) and the duration of the percept (quantified by the time during which the percepts are present) are influenced by dot occlusions, attention to a motion direction, eye movements, or horizontal disparity between motion directions.

5.1 Occlusions

The spatial relationship between moving elements can significantly affect how the visual system integrates different directions of motion. There have been a number of studies showing that additional depth information, such as occlusion and disparity, disambiguate the rotation direction of spheres simulated with parallel projection (Braunstein, Andersen, & Riefer, 1982; Andersen & Braunstein, 1983; Braunstein et al., 1986). From a geometrical point of view, the dots on a horizontal transparent ring that is elevated with respect to the horizontal plane should be perceived as dots lying at different vertical levels on the projection plane (the monitor) depending on the stimulus dimensions and viewing distance (Figure 2a). If this vertical offset is absent, the visual system might not perceive a ring. By varying the vertical spacing between horizontal stacks of dots in our corrugated display, we showed that the two-fronts percept did not result from this cue.

5.2 Attention to a Motion Direction

We reasoned that the attended motion direction would result in an enhanced perception of the surface and could therefore influence the depth ordering. Numerous studies, using single unit recordings or functional imaging, have established that attention shifts can influence neuronal activation levels (for review, see Kastner & Ungerleider, 2000; Treue, 2001). Attention leads to a predominance of responses to attended locations or object features, and a suppression of responses to nonattended locations or features. We showed, however, that attending to motion direction did not influence the percept of either a cylinder or two-fronts.

5.3 Eye Movements

Attention to the motion direction might lead to undesired tracking eye movements. Tracking the dots' path changes the retinal speed patterns; this might be the reason for different percepts. Indeed, fixation had a strong influence on the percept of SFM. Whereas free viewing rendered both cylinder and two-fronts percepts equal, central or eccentric fixation reduced the presence

of the two-fronts percept, leaving the percept of cylinders intact. When observers viewed the stimuli eccentrically, they obtained a strong percept of the cylinder, but the two-fronts perception time was reduced. Notice, however, that even when observers fixated, the two-fronts percept lasted for more than a quarter of the stimulus presentation period.

5.4 Disparity

SFM can provide information about object shape, but near/far relations between the object and the observer are not specified (Wallach & O'Connell, 1953). Binocular disparity, on the other hand, unambiguously specifies near/far relations. Because each of these sources of information specifies what the other lacks, the combination of SFM and binocular disparity provides a more robust representation of an object than the one evoked by using either source alone (Richards, 1985). Braunstein et al. (1986) showed that binocular disparity can disambiguate the sign of depth in computer-generated displays consisting of orthographic projections of texture elements on the surface of rotating spheres. In addition, van Ee and Anderson (2001) demonstrated that there are early interactions between disparity and both motion direction and speed that help in the 3D reconstruction of a scene. We introduced veridical horizontal disparity in the display such that disparity information specified the percept of either a cylinder or the percept of two-fronts. We showed that even for relatively large disparities, both the cylinder and the two-fronts percept ensued when observers attended to that percept.

5.5 Model Predictions and Comparison with Experimental Results

Existing models of SFM perception inspired by the connection structure of MT neurons (e.g., Andersen & Bradley, 1998) predict only two possible stable percepts. This is incompatible with our experimental results, which revealed four stable percepts. In order to explain these two novel percepts, we suggested an alternative model. If the visual system is assumed to be completely unbiased with regard to the depth curvature polarity (convex or concave), our model predicts that all four possible percepts are equally likely to be perceived at any point in time. The experimental data, however, show that most subjects never observe the two-backs percept. In the framework of our model, a bias in the a priori probabilities for assigning convex or concave depth curvature can explain this aspect of the experimental data.

If the a priori probabilities favor the assignment of convex depth curvature, the odds of perceiving the two-backs percept are greatly reduced. Completely eliminating the two-backs percept by this mechanism would, however,

simultaneously, eliminate the cylinder percepts. The absence of a two-backs percept for most observers is therefore probably the result of “high-level” top-down processes similar to those involved in the “hollow mask” illusion. Experiments on this illusion have shown that even when depth cues such as disparity and perspective are present, concave shapes may be perceived as convex. Similar processes may be responsible for the fact that unidirectional flow fields with sinusoidal velocity profiles are predominantly perceived as convex surfaces.

Attentional modulation of the percepts (Figure 3) can be explained by biasing the curvature assignment stage. Because the curvature for the leftward and rightward moving dots is assigned independently, the biasing of convex or concave curvature, too, can be independent for the leftward and rightward moving dots. This allows attention to be focused on any of the four possible percepts.

5.6 Possible Implications for Neurophysiological Studies

Single-unit and neuroimaging experiments have demonstrated that directed spatial attention can increase responses in the human MT+ complex and monkey's area MT (Brefczynski & DeYoe, 1999; Gandhi et al., 1999; Somers et al., 1999; Treue & Maunsell, 1996). It has also been shown that featural attention to motion can selectively increase responses in both human MT+ and monkey MT (Beauchamp et al., 1997; Chawla et al., 1999; Corbetta et al., 1990; O'Craven et al., 1997; Treue & Martinez Trujillo, 1999). Bradley et al. (1998) and Dodd et al. (2001) have shown that MT responses correlate with monkeys' perceptual responses. In the latter studies, monkeys were trained to indicate the motion direction of the perceived front surface, but not to indicate the perceived shape. On the assumption that the stimulus percept was bistable, indication of the motion direction of the perceived front surface was considered to uniquely determine the perceived shape. Given that the monkeys were trained using disparity in the stimulus (rendering an unambiguous depth order), this assumption was probably correct. However, we would like to point out that there is no guarantee it is correct if we take into account the possibility that the two-fronts percept may have been perceived in a number of trials. In this respect, it is worth repeating that cueing to the two-fronts percept is not a prerequisite for perceiving the two-fronts percept. During demonstrations, we encountered a considerable number of observers who perceived both the cylinders and the two-fronts percepts spontaneously, even before they were informed about the possible percepts. Moreover, although for human observers the two-backs percept does not occur frequently, we do not know whether this is true for monkeys; and, therefore, the monkeys might in fact have perceived the two-backs

percept in a number of trials. This is not to say that we should discount the important results reported by Bradley et al. (1998) and Dodd et al. (2001). These results reflect a strong correlation between the neuronal responses and the animals' responses to the motion direction of the perceived front surface. Their results would nevertheless have been more informative if the authors had controlled the attentional state of the monkeys, thereby augmenting the conditions in which the monkeys presumably perceive the cylinder. Further, by controlling the attentional state of the observer, or having the observer report the perceived shape rather than the direction of motion of the nearer surface, it would be possible to determine whether area MT cells respond differently when the subject attends to the cylinder or to the two-fronts. This information might help to reveal whether MT neurons are selective to gradients of motion and depth, or whether MT activity correlates with perceived 3D shape and curvature.

6. Conclusions

We have presented quantitative experimental results and a mechanistic explanation for two so far unreported (but often observed) phenomena that cannot be explained by existing SFM models. (1) The first phenomenon concerns the observation that SFM displays resulting from the parallel projection of a transparent cylinder allow four rather than the two traditionally reported percepts. (2) The second phenomenon concerns the observation that we are able to attentively bias the average period of time during which the different percepts are present.

Although we have gained a wealth of insights from both experimental and theoretical neurophysiological studies in 3D SFM perception, we believe that the studies would be more informative if experimenters had controlled the attentional state of the observers, because it might have been possible for observers to perceive, at will, each of the four percepts described here.

Footnotes

¹ The two novel percepts can be easily seen by color-coding the dot patterns (motion directions). Demonstrations of the stimulus used and effects seen are available on our Web site at <http://www.phys.uu.nl/~vaneer/tube.html>.

² For one of the two observers who were able to create the two-backs percept, the percept of the cylinder was built up faster than the two-fronts percept. The other observer perceived the cylinder for a longer time than the two-fronts, and she perceived the two-fronts for a longer time than the two-backs.

Acknowledgments

We are grateful to Drs. J. J. Koenderink and A. H. Wertheim for seminal discussions that initiated our work. We wish to thank Dr. S. Treue for helpful discussions about both the stimulus and the experimental procedure during his visit to our lab and P. Schiphorst for technical assistance. The authors were supported by the Foundation for Life Sciences (SLW) of the Netherlands Organization for Scientific Research (NWO).

Commercial Relationships: None.

References

- Andersen, G. J., & Braunstein, M. L. (1983). Dynamic occlusion in the perception of rotation in depth. *Perception & Psychophysics*, *34*, 356–362. [PubMed]
- Andersen, R. A., & Bradley, D. C. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Sciences*, *2*, 223–228.
- Beauchamp, M. S., Cox, R. W., & DeYoe, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *Journal of Neurophysiology*, *78*, 516–520. [PubMed]
- Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, *373*, 609–611. [PubMed]
- Bradley, D. C., Chang, G., & Andersen, R. A. (1998). Encoding of 3-D structure-from-motion by primate area MT neurons. *Nature*, *392*, 714–717. [PubMed]
- Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics*, *31*, 261–267. [PubMed]
- Braunstein, M. L., Andersen, G. J., Rouse, M. W., & Tittle, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception & Psychophysics*, *40*, 216–224. [PubMed]
- Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the ‘spotlight’ of visual attention. *Nature Neuroscience*, *2*, 370–374. [PubMed]
- Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*, 671–676. [PubMed]
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248*, 1556–1559. [PubMed]
- DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*, 677–680. [PubMed]
- DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience*, *19*, 1398–1415. [PubMed]
- DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience*, *11*, 219–226. [PubMed]
- Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *Journal of Neuroscience*, *21*, 4809–4821. [PubMed]
- Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *Journal of Neuroscience*, *21*, 4809–4821. [PubMed]
- Dosher, B. A., Landy, M. S., & Sperling, G. (1989). The kinetic depth effect and optic flow: 1. 3-D shape from Fourier motion. *Vision Research*, *29*, 1789–1813. [PubMed]
- Fernandez, J. M., Watson, B., & Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research* *42*, 883–898. [PubMed]
- Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Science U. S. A.*, *96*, 3314–3319. [PubMed]
- Husain, M., Treue, S., & Andersen, R. A. (1989). Surface interpolation in three-dimensional structure-from-motion perception. *Neural Computation*, *1*, 324–333.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341. [PubMed]
- Koene, A., Hol, K., & van Ee, R. (2002). Modeling curvature polarity in multi-stable 3D structure from motion [Abstract]. *Perception*, *31*, S151.

- 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. [PubMed]
- Maunsell, J. H. R., & Van Essen, D. C. (1983a). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3, 2563-2586. [PubMed]
- Maunsell, J. H. R., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49, 1148-1167. [PubMed]
- Merk, I., & Schnakenberg, J. (2002). A stochastic model of multistable visual perception. *Biological Cybernetics*, 86, 111-116. [PubMed]
- Miles, W. R., 1931. Movement interpretations of the silhouette of a revolving fan. *American Journal of Psychology*, 43, 392-405.
- Nawrot, M., & Blake, R. (1991a). The interplay between stereopsis and structure from motion. *Perception & Psychophysics*, 49, 230-244. [PubMed]
- Nawrot, M., & Blake, R. (1991b). A neural-network model of kinetic depth. *Visual Neuroscience*, 6, 219-227. [PubMed]
- Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8, 2201-2211. [PubMed]
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341, 52-54. [PubMed]
- O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18, 591-598. [PubMed]
- Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience*, 8, 4531-4550. [PubMed]
- Qian, N., & Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals: 2. Physiology. *Journal of Neurophysiology*, 14, 7367-7380. [PubMed]
- Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, A2, 343-349. [PubMed]
- Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Science U. S. A.*, 96, 1663-1668. [PubMed]
- Taylor, M. M., & Aldridge, K. D. (1974) Stochastic processes in reversing figure perception. *Perception & Psychophysics*, 16, 9-27.
- Treue, S., Husain, M., & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Research*, 31, 59-75. [PubMed]
- Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. C. (1995). Structure-from-motion: Perceptual evidence for surface interpolation. *Vision Research*, 35, 139-148. [PubMed]
- Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382, 539-541. [PubMed]
- Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575-579. [PubMed]
- Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neuroscience*, 24, 295-300. [PubMed]
- van Ee, R., & Anderson, B. L. (2001). Motion direction, speed, and orientation in binocular matching. *Nature*, 410, 690-694. [PubMed]
- van Ee, R., & Richards, W. (2002). A planar and a volumetric test for stereoaomaly. *Perception*, 31, 51-64. [PubMed]
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45, 205-217.
- 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. [PubMed]