Hidden assumptions in seeing shape from shading and apparent motion

Stuart Anstis

Department of Psychology, York University,
4700 Keele Street, North York, Ontario M3J 1P3, Canada

Introduction

How much of our perception of the world is driven by the immediate local stimulus and how much by stored knowledge and expectations? Hypotheses and expectations can usefully constrain our interpretations of sensory data by capitalising on our knowledge of the real world (e.g. Gregory 1970: Marr 1982: Ramachandran 1990). High-order properties such as depth, transparency and familiarity may help us make an intelligent "top-down" perceptual choice between alternatives in a perceptually ambiguous situation. But, as we shall see, many perceptions can be explained more parsimoniously as "bottom-up" processes driven simply by the luminance distribution in the stimulus. We shall discuss the role of hidden perceptual assumptions in two contentious areas: shape from shading, and ambiguous apparent motion.

Pomerantz and Kubovy (1981) have distinguished between "two different forms of the Pragnanz principle, which can be called the *simplicity principle* and the *likelihood principle*. The first, which is linked closely to the classical Gestalt conception, holds that we organize our percepts so as to minimize their complexity. In information-processing terms, this principle would imply processing small elements of a scene first, then conjoining them into larger clusters, which are then combined into even larger groups until the process reaches a stopping point. Such procedures are known as "bottom-up" procedures.... The second principle, that of likelihood, is definitely not part of the Gestalt heritage but instead may be attributed to Helmholtz. It holds that we organize our percepts so as to perceive the most likely distal stimulus that could have given rise to them. In more modern terms, the likelihood principle would operate via a learned "top-down" process (although evolution could have provided us with a bottom-up process to serve this function)".

Shape from shading

What assumptions are made by the visual system in deriving 3-D shape from shading? It is a fact of physics and geometry that the light reflected from a Lambertian surface depends upon the angle of incidence, so that a curved surface is shaded, and it is a fact of psychology that we can use this shading information to recover 3-D structure. We shall discuss perceptual assumptions that familiar objects such as faces are convex; that light comes from above; that the illuminated side of the object is nearest to the light source; and that light is bright.

1. *Faces are convex.* When we look into a hollow mask it often looks convex and we are simply unable to see it as hollow. This resistance to reversal of depth has traditionally been attributed to familiarity with the shape of objects and the presence of monocular depth cues. Thus, Gregory (1970) attributes it to probability biasing in favour of the likely against the unlikely. He points to the two opposed principles of processing upwards and downwards, the first generating hypotheses which may

be highly unlikely and even clearly impossible, the second offering checks 'downwards' from stored knowledge, and filling gaps which may be fictional and false. But van den Enden and Spekreijse (1989) offer a non-cognitive explanation. A stereoscopic picture of a face offers two kinds of depth cues; binocular disparity, and monocular texture disparity -- gradients of texture, which are geometrically more compressed near the left and right edges of a convex face than they are for a hollow face. They claimed that the real reason that a pseudoscopically viewed face refuses to look concave is that each monocular view contains texture information that provides a strong cue that the face is actually convex. They viewed a convex face through a pseudoscope, and projected 'neutral' texture, which gave no monocular cues, on to the face from a projector near the observer's eyes (Georgeson 1979) Result: the face was correctly perceived as concave. Deutsch and Ramachandran (1990) and Peli (1990), however, both point out that this projected texture adds rich binocular disparity cues, which suffice to explain van den Enden and Spekreijse's results. Furthermore, the texture account does not explain why an actual hollow face, viewed with both eyes, looks convex. However, Deutsch et al. and Peli concur that the perceived depth depends upon the cues present in the stimulus, and no cognitive factors need be involved.

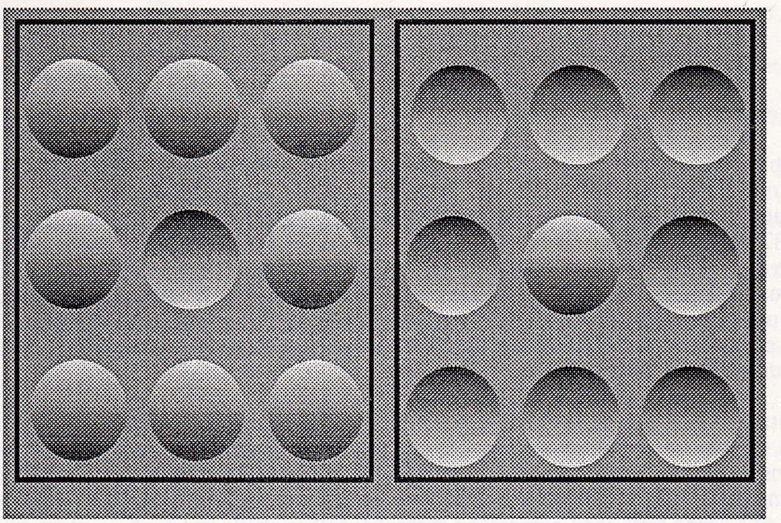

Figure 1.

2. *Light comes from above.* Fig. 1 b is simply Fig. 1 a turned upside down. As a result the picture appears to be reversed in depth, with the bumps becoming hollows and vice versa. In the absence of explicit lighting cues the visual system assumes that the light comes from the top of the picture. The assumption that laight comes from above rather than any other direction does not simplify our perceptions in any way but it does match them better to our sunlit physical world in which the light really is more likely to come from above, so we are applying a perceptual constraint derived from our knowledge of the physical world. This is a prime example of the likelihood principle.

Ian Howard (personal communication) has noted that if you bend over and look between your legs at Fig. 1, the light is assumed to come from above in retinal, not gravitational coordinates, as if the light source were assumed to be stuck to one's forehead. This is not a highly intelligent perceptual assumption.

There is some evidence that the constraint is learned by experience. Hess (1972) reared chicks in special cages that were lit from below through the floor, so that the grains they ate were illuminated from below. He then exposed them to a pair of photographs of grain. In one photograph the illumination came from above: in the other, from below. The chicks pecked at the grains illuminated from below. Once learned in early life, however, such a constraint could arguably become hard-wired.

3. *The illuminated side of object is nearest the light source.* This sounds so trivially obvious as to be barely worth saying. For myself I only became aware that I make this assumption when it appeared to be violated by Brian Rogers' New Moon Illusion. This phenomenon was first pointed out to me (on Moon Drive in Toronto) by Brian Rogers. (It is called the New Moon Illusion because it was described more recently than the Old Moon Illusion in which the moon looks larger when it is near the horizon. It is seen best when the moon is about half full, and it has nothing to do with the new moon, else it would be called the New Moon Illusion.)

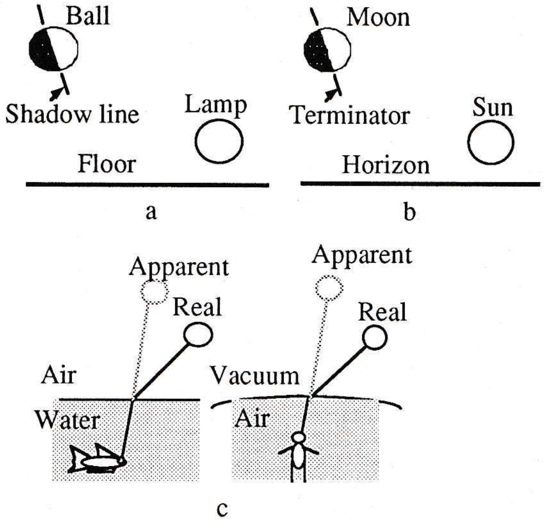

Figure 2 **a,** Obviously this picture is wrong. The shadow line on the ball is tilted to about 11 o'clock, but it should be tilted to 1 o'clock since the lamp is lower than the ball.
b, Brian Rogers' New Moon Illusion. The terminator line on the half-full moon looks tilted to 11 o'clock even though the sun looks lower in the sky than the moon.
c, The new moon illusion is not caused by atmospheric refraction, which makes the sun look *higher* in the sky, not lower, than it really is.

In Fig. 2a the lamp is below the sphere, yet the edge of the shading on the sphere is inclined to the left of vertical at about '11 o'clock' indicating that the illuminating lamp must be higher than the sphere. Anyone can see that there is something wrong and the picture is incorrectly drawn. Fig. 2b is similar but the sphere has been replaced by the moon and the lamp by the sun. The shadow line on the moon, known to astronomers as the terminator, is tilted at about 11 o'clock, so one predicts that the sun ought to be higher than the moon. On looking over one's shoulder at the sun, however, one finds that the sun is lower in the sky than the moon. This appears to contravene the laws of physics, and it certainly contravenes one's perceptual assumptions. I recently attended an evening party when the sun was setting but the moon was already high in the sky, and I asked several fellow scientists to explain this (with hypotheses based upon explicit assumptions about the physical world, as opposed to any implicit perceptual assumptions underlying the illusion itself.)

A mathematician, perhaps inspired by Ptolemy, suggested that refraction makes the sun look lower than it really is. This assumption is false because atmospheric refraction actually bends the sun's rays so as to make the sun look *higher* than it really is, and keeps the sun visible for a few minutes after it has sunk below the horizon (Fig. 2c). So refraction cannot explain the new moon illusion since it predicts an effect in the wrong direction.

An astronomer suggested that the orbits of the sun, earth and moon lie in different planes. This assumption is true, but irrelevant. The new moon illusion would still be visible if the scene were viewed in a single brief flash, and motions of the heavenly bodies make no difference.

A distinguished physicist suggested, admittedly toward the end of the party, that the sun, moon and observer do not lie in the same plane. This assumption is false, since as his wife (the mathematician) reminded him, any three objects always lie in the same plane.

Michael Swanston (personal communication) has convincingly argued that the sun and moon look about equidistant from the observer, although the sun is really 400 times as far away as the moon. This misperception of the sun's position, based on a gross underestimate of its distance, leads to the illusion. Probably the old nor the new moon illusions are both due to the misperception of distances from the observer -- of the moon itself for the old moon illusion and of the sun for the new moon illusion.

4. *Light is bright.* Again this sounds too obvious to be worth saying. What else could light be but bright? Actually it is counterfactual, but not meaningless, to suggest that light could be dark. One could simulate a lamp with a spray can spraying out white paint on to black objects. When the paint dried it would give a fossilised rendering of illuminated objects. Black paint sprayed on to white objects does not give a rendering of any physical light source because there is no "black light" in nature. Of course, in principle one could construct an *internal* world of black light and white shadows, not by changing the external physical universe but by re-wiring the visual system so that highly reflective objects produced the sensation that we now label "black", and unlit space would produce the sensation that we now label "white". But this is simply to restate the old speculation that although we agree to call a ripe cherry "red" and a leaf "green", it may be that my "green" sensation would look like your "red" sensation if they were somehow both fed into the same brain. More to the point, we can interchange black and white *experimentally* by viewing the world through a video link which converts the picture electronically into a photographic negative. This makes it hard to distinguish bumps from hollows and impossible to recognise faces of celebrities, especially if the portraits are high-contrast "lith" photographs (Phillips 1972). The brightness reversal disrupts the shape from shading which permits us to recognise facial features. Patrick Cavanagh was able to establish, after two years' work, that shadows look like shadows only if they are darker than illuminated regions (Cavanagh and Leclerc 1989).

Our ability to decode shadows may have a learned component (Hess 1972). Although naive observers cannot recognise famous faces in negative, there was a time when people who worked in television newsrooms routinely acquired this unusual skill. In the early days of television, newsreels were shot on cine film. The negative film stock was developed and a positive print made for broadcasting. It was soon realised that valuable time could be saved by broadcasting the negative film directly and reversing the brightnesses electronically. The film editors were called upon to edit the films while these were still in negative, and they rapidly learned to recognise world leaders even in negative. Suneeti Kaushal and I are planning to investigate long-term adaptation to a negative visual world. Subjects will wear a helmet-mounted stereo display in which two small TV cameras on the front of the helmet feed into two miniature TV screens which the subject views, one with each eye, through suitable magnifying lenses. The TV pictures will be electronically reversed in brightness.

Apparent motion

As a microcosm of perceptual processes we used ambiguous apparent motion stimuli in which two shapes abruptly exchange positions. One can perceive the two objects as moving past each other in opposite directions, or one object can seem to move while the other does not, or each object can change shape without shifting position. Which of these percepts occurs is strongly influenced by stimulus and observer variables, and we shall discuss the role of "bottom-up" low-level processes and "top-down" assumptions, expectations, or learned constraints about the physical world.

Braddick (1974) proposed a distinction between short-range and long-range processes in motion perception. The short-range process is thought to occur early in the visual system and has been identified with directionally selective neurons in the visual cortex that operate passively and in parallel over the whole visual field. It operates over short distances (<15 min in the fovea) and brief durations (<100 ms), and adaptation of the short-range system underlies the motion aftereffect. Its inputs come exclusively from stimuli that are defined by luminance. The long-range process, on the other hand, is thought to occur at later stage of processing with properties more resembling cognitive or interpretative processes than the responses of single neurons. It operates over longer distances and times than the short-range process, and can accept non-Fourier stimuli as inputs, for example patches that are defined by texture, cyclopean depth, or short-range motion (Cavanagh 1989). The notion of short and long range motion processes has been reviewed by Braddick (1980) and Anstis (1980). Cavanagh and Mather (1989) have published a highly critical review of these concepts, but they certainly still have heuristic value. Motion perception has been reviewed by Nakayama (1985), Borst and Egelhaaf (1989) and Sekuler *et al* (1990), and Newsome *et al* (1989) have directly compared motion perception by an alert monkey with the performance of its neural motion detectors.

It is plausible to equate the short-range process with bottom-up processes and the long-range with top-down. Marr and Ullman (1981) proposed a computational model of the short-range process and suggested that there are two types of computational tasks associated with motion perception, tasks of separation and tasks of integration. Tasks of separation can be solved in principle by using only instantaneous measurements such as position and its time derivatives in the image. This includes such tasks as motion segregation, and could probably be handled by short-range processes. Tasks of integration cannot be solved using only instantaneous measurements but require the combination of information over time. This includes such tasks as the recovery of 3-D structure from motion (Ullman 1979) would probably require long-range processes. Ullman (1979) proposed a computational model of the long-range process, which involves computing similarities across successive time frames to match up correspondence tokens, and then does a cost-benefit analysis which essentially minimises the total path lengths of all motions in the visual field.

A parsimonious motion system would operate only on the stimulus luminance. We shall discuss luminance, illusory brightness, and texture segregation. A more intelligent motion system would first analyse a scene for depth, using such cues as perspective, shape from shading, and occlusion, and then use 3-D objects as correspondence tokens. Examples are given below.

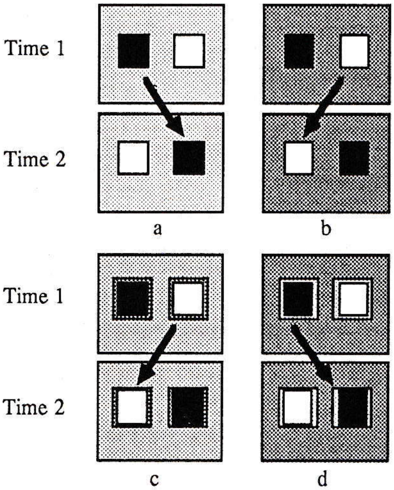

Figure 3 Apparent motion when a black and white square suddenly change places (Anstis and Mather, 1985).
a, On a light surround the black square appears to move.
b, On a dark surround the white square appears to move. Thus apparent motion is attributed to the square with the higher luminance contrast.
c, When the surround is light but each square is surrounded with a dark picture frame, the white square usually appears to move.
d, When the surround is dark but each square is surrounded with a light picture frame, the black square usually appears to move. Thus luminance contrast is assessed by local mechanisms.

High-contrast objects appear to move

Luminance. Fig. 3a,b shows a black square and a white square side by side which instantaneously change places, so that the black square suddenly becomes white and at the same instant the white square becomes black. What will one perceive? Does one see two squares flickering in place, or does the white square jump one way, or does the black square jump the other way? George Mather and I (1985) found that the answer depends upon the luminance of the background. On a light surround the black square is seen as jumping (Fig. 3a), and on a dark surround the white square is seen as jumping (Fig. 3b). The square that differed most from the surround, in other words the square with the higher contrast, was perceived as jumping.

The observers adjusted the luminance of the background until they saw either the two squares move equally frequently or both squares move at once. This mid-grey indifference point was equivalent to a paint mixed from equal parts of black and white paint, so it was the arithmetic, not the geometric mean of black and white. This shows that the motion system operates linearly on luminances, not upon log luminance. More recently, however, Shioiri, Cavanagh and Favreau (1989) have found a slight non-linearity - less marked than a logarithmic function -- which is consistent with the compresson of retinal cone responses reported by Boynton and Whitten (1970). Shioiri *et al* (1989) have also used this technique to study the linearity of intensity coding along the three cardinal axes of color space. Two bars of two colors (C1, C2) were alternated to produce apparent motion. The background color was a mixture of C1 and C2. If the background color was closer to C1, only the C2 bar appeared to jump. If the color different in a linear cone activation space controls this apparent motion, the setting should be midway between C1 and C2, no matter which pair of colors is chosen for C1 and C2. None of the cardinal axes (achromatic, R-G cone, or B axes) showed a linear response, but the non-linearity was less extreme than a logarithmic function.

The surround can be partitioned into a large background area and a small area like a picture frame adjacent to the test squares (Fig. 3c, d). When these areas are pitted against each other by setting them to different luminances, the picture frame luminance has a much stronger influence than the remote surround on the perceived direction of motion.

Illusory brightness. Fig. 4a shows two pieces of black/grey checkerboard, each moving back and forth through the diameter of one checkerboard square. The two moving checkerboards do not interact. Now a surround is added -- a large stationary black-white checkerboard, positioned so that each small grey checkerboard replaces alternately some black squares and some white squares of the surround (Fig. 4b). White (1979, 1981: White and White 1985) showed that grey squares that replace white squares in a checkerboard look appreciably darker than when they replace black squares, as Fig. 4b shows. (White used gratings, not checkerboards, but the principle is the same). Now the two regions interact strongly, and a single region is seen as jumping back and forth through a dozen square-diameters. The illusory induced brightness controls the apparent motion just as effectively as the physical luminance did in Fig. 3.

White's phenomenon has attracted considerable interest because it cannot readily be classified as either simultaneous contrast or as brightness assimilation (Hamada, 1984: Foley and McCourt, 1985: Moulden and Kingdom, 1989: Zaidi, 1989). The grey squares that replace white squares are bordered by black squares, and from simultaneous contrast grounds would be predicted to appear lighter than the grey squares that replace black squares, the exact opposite of what is found. Moulden and Kingdom (1989) attribute White's effect to two mechanisms, one a local concentric spatial filter operating at the corners of the grey test regions, the other a spatially extensive filter operating along the long bars of the grating (or presumably along diagonal rows of checkerboard squares).

To perceive a single very large jump instead of two small jumps seems to violate principles of simplicity and likelihood. We believe this is a bottom-up process that accepts only low-level luminance cues. First, a comparison of the left hand regions of Fig. 4 at Times 1 and 2 shows that the edges of the checkerboard squares reverse their luminance polarities in the transition from Time 1 to Time 2. This will militate strongly against seeing the small local motions (Anstis 1970: Anstis and Rogers 1975: Anstis and Mather 1985). Second, adding the background checkerboard changes the

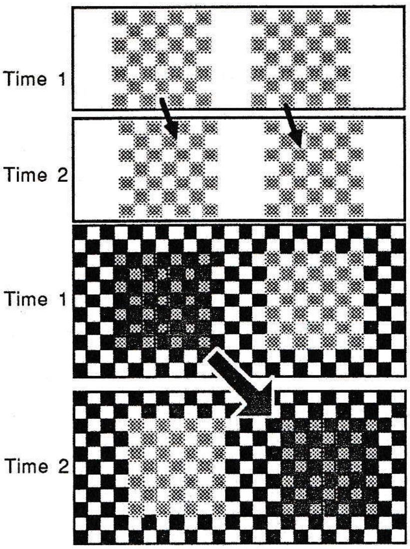

Figure 4 Illusory lightness can mediate crossover motion.
a, two grey checkerboards each move through the diameter of one checkerboard square. Each checkerboard appears to move locally and independently (grey arrows).
b, a stationary black/white checkerboard surround is added. In the left-hand region at Time 1, the white checkerboard squares have been replaced by grey squares. This region looks darker than the right-hand region in which the black checkerboard squares have been replaced by grey squares (White, 1979). At Time 2 the two regions interchange. Result: the higher contrast region, here the apparently dark one, appears to jump across to the right (black arrow).

whole luminance distribution, especially at low spatial frequencies; spatially integrating Fig. 4 by blurring or squinting at it will reveal the left hand region to have an appreciably lower space-averaged luminance than of the right-hand region, so a large-scale visual filter that integrated the stimulus neurally would pick up a Fourier energy motion signal. Of such scale effects the Gestalt psychologists who pioneered the simplicity principle knew nothing.

Thus, the responses to the real and illusory brightnesses in the stimuli of Figs. 3 and 4 can be explained by a low-level process sensitive to motion energy. The two squares in Fig. 3a can be compared to two superimposed gratings of the same spatial frequency moving in opposite directions. If the two gratings are of equal contrast the outcome is a stationary counterphase flickering grating, but if one grating, say the one that moves to the right, is of higher contrast then the combined stimulus contains more motion energy to the right and is perceived as moving to the right.

Texture segregation. We can generalize these results to squares defined by texture, not by luminance, where there is no Fourier energy moving predominantly in one direction, yet the direction of perceived motion is controlled by the perceptual *salience* of the textured squares. Fig. 5 shows two textured squares consisting respectively of coarse and fine random dots, on a surround of very fine random dots (Fig. 5a) and again on a surround of very coarse dots (Fig. 5b). The dot diameters in the fine surround, the two squares, and the coarse surround, are in the ratios 1: 2: 4: 8. Although all four textures have the same space-averaged luminance, it is easy to segregate the two squares perceptually from the surrounds. However, the finer textured square stands out as more salient against the coarse surround, and the coarser textured square stands out better against the fine surround. This texture salience controls the apparent motion, although less compellingly than luminance did in Fig. 3. The two squares suddenly exchanged places and observers were asked to report whether they saw the fine texture jumping to the left or the coarse texture jumping to the right. We found that the answer depended upon the texture of the surround. When the background texture was coarser than the coarse square, the finely textured square was perceived in apparent motion, and when the background texture was finer than the fine square, the coarsely textured square was perceived in apparent motion. The square that differed most from the surround was seen in motion.

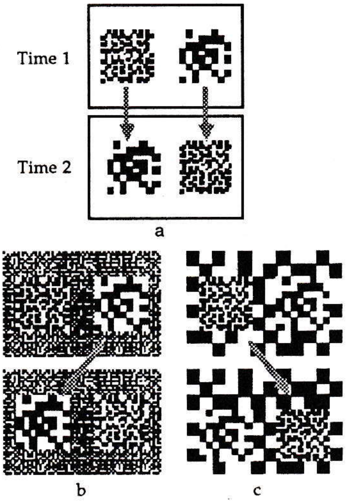

Figure 5 Perceptual salience, based on texture discrimination without luminance cues, can mediate crossover motion.
a, two patches of random-dot texture, one fine and one coarse, exchange places. Apparent motion is ambiguous.
b, when the stimulus in **a** is superimposed on a very fine textured background, the coarse texture is more salient and is seen in apparent motion (arrow).
c, when the stimulus in **a** is superimposed on a very coarse textured background, the fine texture is more salient and is seen in apparent motion (arrow).

To summarise, our crossover effects appear to operate at a fairly low level, after the level of luminance discrimination or of texture segregation. They require no simplicity or likelihood assumptions by the visual system. Motion perception can be based upon stimulus luminance (Figs 3, 4), or at a higher visual level upon texture-based non-Fourier stimuli (Fig. 5), which admittedly provide somewhat less compelling impressions of movement (Chubb and Sperling 1988). We conclude that stationary regions can first be defined by any visual cue such as luminance, depth, texture and so on, and then displacements of such regions can then give rise to a long-range motion percept. Cavanagh (1989) reviews many examples of such inter-attribute apparent motion.

Near objects appear to move

Perspective. Fig. 6 shows a perspective sketch of a protruding square slab next to a square recess. When the slab and the recess abruptly exchanged places, subjects reported that the slab, not the hole, moved. When the color of the bottom of the recess was made the same as the top of the slab, the display now looked like two buttons being pushed in alternation and subjects now reported motion in depth, along the line of sight and at right angles to the previous motion. (The stimuli used were these actual sketches, not real 3-D objects). These are examples of intelligent processes in long-range motion perception.

Shape from shading. Fig. 7 shows the familiar Ternus (1926) configuration for apparent motion. Three spots jump back and forth between positions a,b,c and time 1 and positions b', c', d' at time 2. The percept depends upon the timing. When there is no interstimulus interval (ISI) the two central spots are always visible and look stationary, and subjects report 'element motion' in which one spot jumps from end to end (Fig. 7a). When there is an ISI the central spots flash on and off and subjects report 'group motion' in which three spots jump back and forth together (Fig. 7b)(Pantle and Picciano 1976). Group motion is also seen when the stimuli are presented dichoptically. Braddick and Adlard

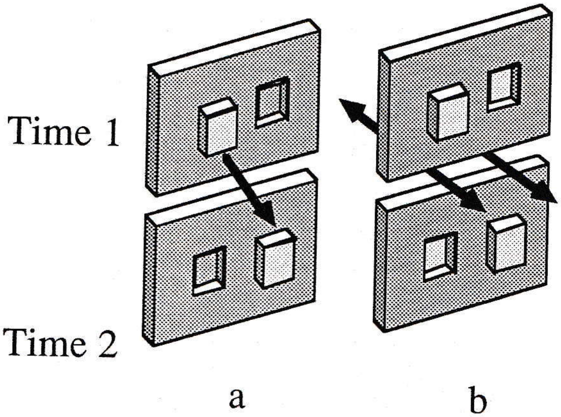

Figure 6 Apparent motion is attributed to the nearer object.
a, When these two perspective sketches are superimposed and exposed in alternation, the protruding slab is seen in motion (arrow) between two stationary holes, moving in the plane of the wall.
b, When the color of the bottom of the recess is made the same as the top of the slab, the display looks like two buttons being pushed in alternation and the motion is now in depth, along the line of sight (arrows).

(1978) found that when they restricted the dichoptic motion to the two central spots, group motion was often reported, but when they restricted the dichoptic motion to the two outer spots subjects nearly always reported element motion. Thus the effect of dichoptic presentation in reducing element movement depended, oddly enough, not on dichoptic presentation of the element that apparently moved, but of the elements that appeared stationary. They concluded that element motion is mediated by a low level short-range process (Braddick 1974) --not by sensing that the end element is jumping back and forth, but by sensing that the two central spots are stationary. They attributed group motion to a higher level, more interpretative long-range motion process. For a differing view see Cavanagh and Mather (1989).

Now let us give the disks some apparent depth by means of shape from shading (see Ramachandran 1988). The flat grey disks of Fig. 7 a, b are replaced in Fig. 7 c, d by shaded disks that look like saucers, and the empty spaces by shaded disks that look like bumps. When there is no ISI, instead of seeing element motion most observers report that the disks at the left and right ends of the display are flipping up and down without changing their position, driven by the "dumb" local luminance cues. However, when there is an ISI, instead of seeing a group of three saucers moving to the right, most observers report a single bump jumping to the left between the endmost of four empty saucers. So the long range motion is controlled by "smart" cues of perceived depth. It seems that the motion tokens for long range motion, but not for short range motion, are derived after shape from shading has been computed.

Covering and occlusion. Sigman and Rock (1974) explored the role of occlusion in apparent motion. They propose that the perception of apparent motion can be the outcome of an intelligent problem-solving process. They exposed two stationary spots a and b in alternation, by moving an opaque rectangle back and forth, alternately covering and uncovering the two spots at a tempo that ought to give good apparent motion. As far as other theories of apparent motion are concerned, there is no reason why these conditions should not produce an impression of a and b moving. But from the standpoint of problem-solving theory, the moving rectangle provided an explicable basis for the appearance and disappearance of a and b, namely that they are there all the the time but are undergoing covering and uncovering. This is what the observers reported; they rarely reported apparent motion. However, if the rectangles were drawn so as to look transparent they did not look capable of covering anything, so it was no longer a fitting or intelligent colution to perceive a and b as two permanently present dots that were simply undergoing covering and uncovering. In this condition, subjects again reported apparent motion (Rock, 1983).