17

Movies in the Mind's Eye

Julian Hochberg
and Virginia Brooks

Most writers on film, and most filmmakers, need no science. But any serious discussion of whether the medium was used effectively or artistically in any instance requires some understanding of how we perceive and remember moving pictures, and that must derive from research: introspection will not serve. Scattered aspects of cognitive science have begun to appear, therefore, in recent writings on film.[1] On the other side of the aisle, students of perception and visual memory cannot afford to ignore moving pictures, but until recently they have mostly confined their attention to low-level motion phenomena, as have introductory film texts. The latter, if they write at all about perception, still proclaim that we perceive motion from successive still frames because of (heavens protect us!) "persistence of vision."[2] In any case, stroboscopic motion is only a small part of the perception of visual events, which is what film is about.

In this essay we reexamine the cognitive systems that contribute to the visually informative and artistically important characteristics of film and tape, trying to keep both the science and the art in view.

Depicting Events in Moving Pictures

There are three steps to the depiction of events in moving pictures: low-level vision, relational parsing, and action schemas. We will look at each separately.

Movement as Primitive Sensory Response

A continuous motion in the world is, of course, captured by successive displaced images on film (or their video equivalent). For most events, these displacements are small, and within the range of the low-level sensory receptors of the visual system; these respond identically to the small displacements on the screen and to the differences provided from one moment to the next by

smooth physical motion in the world. Recent studies of this system offer an increasingly important window on the underlying neurophysiology, and reveal some surprising phenomena (for example, reversals of direction).[3] This low-level response is preemptive and "unintelligent." Notably, it occurs between nearby successive contours regardless of what objects they belong to (see, for example, Fig. 17.3A), thereby causing many "bad cuts" (which result from this fundamental displacement-detection mechanism, rather than some violation of cinematic "grammaticality"). Many techniques aimed at achieving "seamless" editing work by avoiding such unwanted apparent motion between noncorresponding objects. In the time of Eisenstein, however, and especially in the New Wave, such visual jolts became desirable, although (or because) they slow the viewer's comprehension and make the medium itself more intrusive.

Although vision psychologists and neurophysiologists sometimes write as though these low-level mechanisms account "directly" for perceiving motion, which would, if it were true, make it easy to explain and predict what people will see, it is simply not true: we usually perceive movements very different from the displacements in the eye or on the screen. Indeed, were it not for these differences, films as we know them would not be possible, as we see next.

Framework-Relative Paths of Motion

We perceive (approximately) the framework-relative paths of motion, and not the displacements on the screen which determine low-level motion.

An object may be ***perfectly stationary*** on the screen and yet it will appear irresistibly to move if given a moving framework (Fig. 17.1Ai) or background (Fig. 17.1Aii), and the actual motions of the frameworks or backgrounds themselves are often not noticeable. This is part of a rich body of phenomena known as ***induced motion***.[4] Something akin to Fig. 17.1Ai happens outside of the laboratory when the eye tracks a moving object in a ***pursuit movement*** or when a camera acts similarly in a ***pan*** or ***track*** shot (Fig. 17.1Aiii).[5] Thanks to this phenomenon, a continuous motion can be presented over a space *s'* that may be many times larger than the screen, in the same way that the movements of the viewer's head and eyes provide a wider prospect than the limits of gaze within any glance. The screen would be stage-bound, were it not for this resource. Similarly, the parts of an object or group of objects moving in one direction on the screen may instead be irresistibly seen as moving in another (Fig. 17.1Bi, 17.1Bii).

Neither of these demonstrations is only about dots in the laboratory. The phenomenon of Fig. 17.1A reappears in Fig. 17.1C: The rightward movement of a dancer across the screen (M2) is lost if filmed in limbo (with little

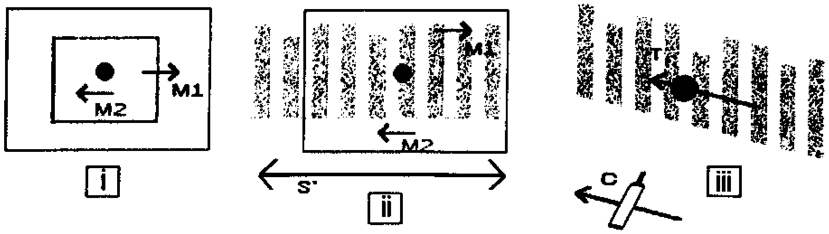

Fig. 17.1A. *i.* Although it is perfectly stationary on the screen (indicated by the outermost rectangle), the black dot appears very strongly to move leftward (M2) if a large object, the background, or the framework actually moves rightward (M1). *ii.* If the background continues to enter on the left edge of the screen and scroll off the right edge, for a total distance (s') that is larger than the screen, the induced movement (M2) can continue as long as the filmmaker wishes it to. *iii.* Such scenes are usually made by moving the camera, C, in synch with the target, T, in a tracking or pan shot.

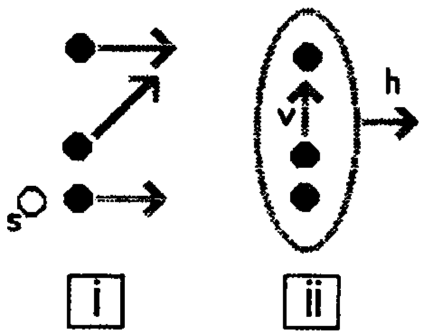

Fig. 17.1B. With three dots moving on the screen, as shown by the arrows in *i,* a very different event is perceived, especially if only the dots are visible (that is, white dots in limbo). *ii.* the center dot is seen in vertical motion (arrow *v*) between the two end dots. If the surroundings are not completely featureless (that is, if the edges of the projection screen are dimly discernible, or there is a stationary dot, like that at *s*), the entire set of dots may be detected as moving horizontally (arrow *h*).

or no framework), as in Fig. 17.1Ci. Indeed, the movement is perceived as ***leftward*** (M4) if the corps, as background, moves rightward on the screen (Fig. 17.1Cii). The phenomenon of Fig. 17.1B reappears in Fig. 17.1D: At (i), we show the first and last frames in a dance movement across the stage; without a strong background, the curve that the lifted dancer's arm describes across the screen in Fig. 17.1Di is not perceived, and only the arm's movement relative to framework provided by the moving body is perceived. Intermediate frames are shown at (ii), in which the curved movement is essentially lost in any case because the dancers are kept centered by a pan shot.

In analyzing a narrative fiction film, such a microscopic approach makes it difficult to get back to the macroscopic content; in films of dance, and other

choreographed spectacles, and indeed in any film in which the visual content is defined by the subject, the camera and the editing are at once relatively open to visual study and critical to an appreciation of the film, such detailed study of the movements seems as important to an understanding of the film as it is to our understanding of cognitive process.

And that leads us to ask: If the movements we perceive may thus be very different from those we can measure on the screen, how can we say in advance what they will be? As part of a more general solution, it has been argued that the background is specified as stationary in the world because it remains otherwise invariant while moving as a whole, and that the relative motion between object and surround is attributed to the eye or camera. If that is true, then a layout of space equal to $s' = m1 \times t$, where s', m are as labeled in Fig. 17.1A and t is m's duration, is also specified, as is the apparent velocity ***m2*** of the target (although really stationary on the screen). A similar explanation has been offered for Fig. 17.1Bii, in terms of the vector that remains after subtracting the component, *h,* that is shared by all the moving dots.[6]

This sounds like an automatic prescription, but it cannot serve as such. For one thing, there is always a stationary visible framework—the edge of the screen—which by the preceding analysis should restrict our perceived motions to the physical displacements within that framework. But we have seen that that does not happen. Indeed, under normal viewing conditions (that is, not in limbo), both on stage and on the screen, the viewer sees the graceful curve as well as the body-relative movement (in Fig. 17.1Di), in some manner hard to describe and even harder to measure. The distinction between these two motions is analogous to the distinction between the 3-D spatial layout represented in a picture, and its pictorial composition: the former is view-

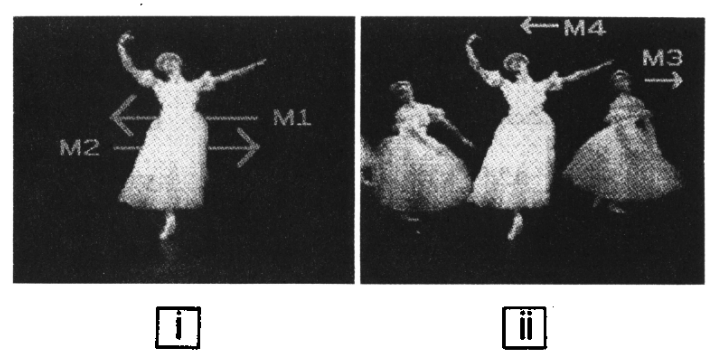

Fig. 17.1C. *i.* A dancer or actor actually moving rightward in space (M2) against a weak or featureless background, but tracked by the camera so as to remain stationary on the screen, will in fact appear stationary. *ii.* The same dancer, against a background (here, the corps) with a net rightward screen movement (that is, M3 > M4), will appear to move leftward (M4), against its true direction of movement in space.

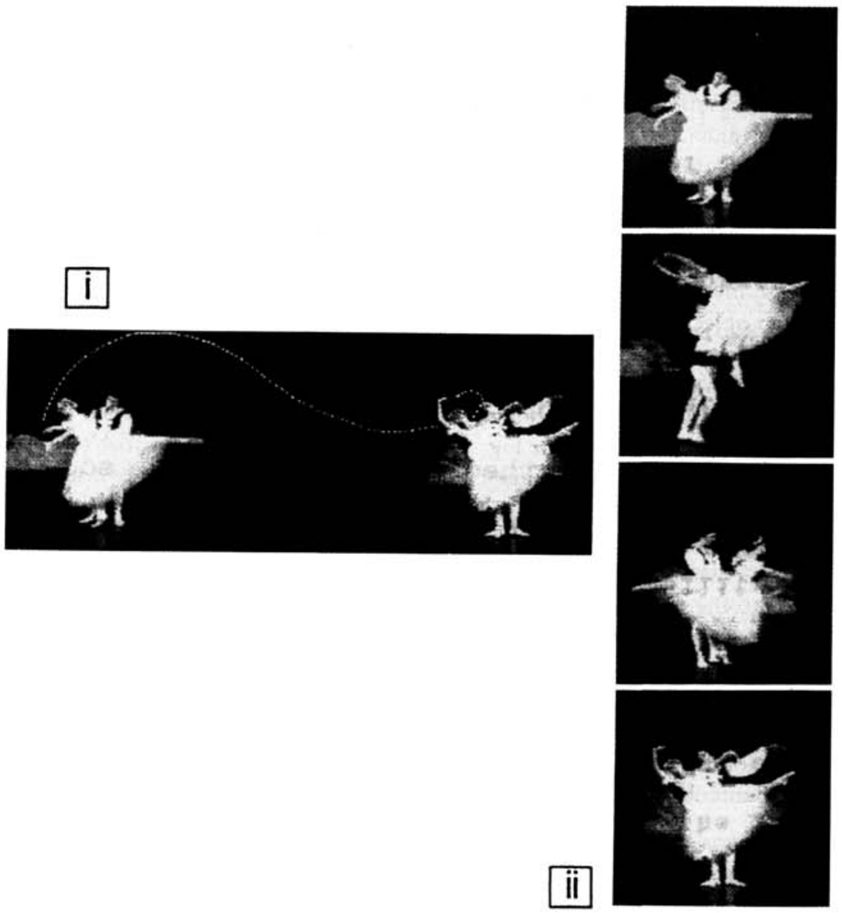

Fig. 17.1D. As in Fig. 17.1B: Although the dancer's hand traces the curve across the screen as drawn in *i*, that curve is not seen if there is no background; the hand movements relative to the body (intermediate frames are shown in *ii*) are seen instead.

point-independent, whereas the latter, which greatly affects the feeling and aesthetics of the view, depends very much on the viewpoint of eye or camera.

Which movement actually predominates in any filmed passage seriously affects the aesthetics of the passage. The balance depends, among other things, on where the viewer attends. For example, a stationary background spot aligned as at *s* with the diagonal motion in Fig. 17.1Bi will, if the viewer stares at it, make the diagonal movement much more visible.[7] The filmmaker learns by trial and error—not by any principles in the production handbook—how to obtain the desired effect.

How can we know in advance what will occur and why?

A set of perceptual theories, each of which uses quite different terms (unconscious inference, perceptual logic, internalization of the laws of physics, and ecological realism), are all versions of Helmholtz's likelihood principle[8]: ***we perceive that which would in our normal life most likely have produced the effective sensory stimulation we have received.***

That principle must surely be at least approximately true, or we could not survive. It is probably a good source of intuitions for filmmakers and other visual artists. But within this set of theories lie two extremely different subsets. Theories of the first class assert that some internal mental representation of the event or scene is formed within the viewer's mind, in response to the movements on screen and in the eye; such theories are frequently accused of being uneconomical and mentalistic (a term with pejorative overtones). To us, a more damaging aspect is that such theories are not predictive, because they have not addressed the nature of the representations they postulate in general; or the nature of what those representations are like in the case of moving pictures in particular.

A second kind of theory of visual perception and cognition seems much more hard-headed, specific, and based in the real world (that is, more "behaviorist"), referring only to the information offered the viewer, and to what that information specifies about the world: the extent of surface defined in Fig. 17.1Bii, for example, or when and where we will make contact with an approaching surface or a thrown ball. In this class of theory, it may be thought that we as scientists or as moving picture craftsmen need only to know the relevant principles of physics to know what people will perceive.[9]

Don't count on it. Whether or not such nonmentalistic or physics-specific theories will ever be good or useful replacements for their mentalistic counterparts for any purposes (and we do not think they will), they fail us here for two reasons that should tell us much about visual cognition quite generally. First, they work by defining motion in a space-time coordinate system, with time as a dimension. It is easy to do this when discussing physical events that can be predicted by physical laws, such as the distance s' traversed in Figure 17.1Aii. But after some brief time, the space and movement of film gone by can exist only in a fallible viewer's limited working memory, and they do not remain unchanged somewhere simply because they had been "specified" by information that has flowed past on the screen. Second, motions that follow simple physical laws that we can hope to formulate do not make up much of moving pictures in any case.

Consider the first point, that is, the role of perceptual memory as opposed to physical specification. In situations like Fig. 17.1Aii, in which a motion $m1$ brings some offscreen surface into view, viewers can judge when remembered objects, currently beyond the screen, will come into sight. Yet experimental measures show that such space in the mind's eye is compressed when out of sight, that is, $s' < m \times t$.[10] Studies in which an object moving at some velocity goes out of view for some brief period t, and the viewer reports whether it was early or late when it reappeared (like testing the timeliness of the return from a cutaway, or of a strolling actor's reappearance from behind an occluding object), found the ability lost within 1,200 msec.[11] In describing what the viewer gains from a moving picture, both of these kinds of study reveal mental representations (effective memories) of motion and of constructed space, rather than merely the specified physical variables. The former is thus not merely a more mystical phrasing of the latter: they have different properties. But it is even more important to us that represented movements and extents do not long outlast their presentation on the screen.

This means that, very shortly after it has occurred, the representation of an event, or of a part of an event, is different from the perception obtained during the event itself. Some specific physical information about space and time is lost with time. We assume that such losses occur as well after each change in direction or speed, or after each cutaway or change in scene. That is, unless the viewer has available some mental structure or schematic event into which the segments take their place, ***and from which they can be regenerated when needed,*** the continual movement in space becomes indeterminate in memory.

This, in turn, is what gives the second point—that most moving pictures are not assemblies of simple physical trajectories—its theoretical and practical importance.

Beyond Physical Trajectories: Represented Goals and Intentions

The fact is that we must parse most of the motion patterns we encounter in terms of purposeful acts, not in terms of physically specifiable trajectories. They are not the same, in that the identical act can be expressed and represented by very different physical motions. In Fritz Heider and Marianne Simmel's film (Fig. 17.2), a triad of geometric figures interact in a rich social narrative, across more different trajectories than viewers can remember if they try to do so in terms of geometric paths, but which they can remember far better as purposive acts within a story structure.[12]

The story structure encodes the too-numerous trajectories as a smaller number of more distinctive and familiar, purposeful actions. These are the same units by which the motions of people and animals achieve recognizable organization (and which are perhaps learned by infants even before they learn simpler physical motion). All of the unrelated movements are immediately meaningful if the viewer has undertaken to construe them as purposeful and expressive actions.

The Heider and Simmel film itself was an explicit request for a new look at social psychology, at narrative, and at visual cognition. Since it appeared in 1944, social psychologists have shown that viewers agree consistently about the ***breakpoints*** (the bounds of staged purposive actions); they have also shown that sequences assembled from breakpoints alone are better comprehended than those assembled from intermediate stills. (For this reason, films seem a natural joint venture for film theory and social psychology.)[13]

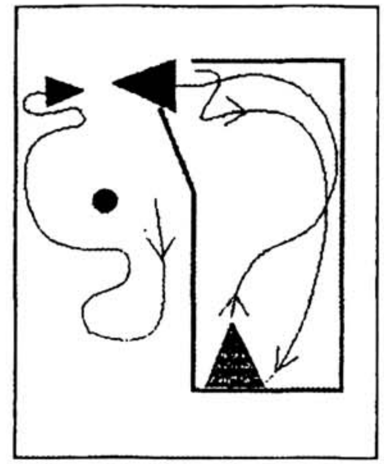

Fig. 17.2. Motion paths of an animated story film by Heider and Simmel. Cast: small circle, large and small triangles (the gray triangle marks a particularly lengthy pause).

The physically defined (nonmentalistic) approaches to visual cognition say nothing about such animate events. But even if they did, we should have to abandon these approaches, designed to make mentalistic accounts unnecessary, when we consider how cuts are used in moving pictures.

Cuts, Story Structure, and the Nature of Movies in the Mind's Eye

It is not clear that any of the more mentalistic likelihood theories can be put in working order as theories about the mind. In general, although after the fact they can plausibly explain why something was seen as it was, they cannot successfully say ***in advance*** what movement will be perceived, or what the mental representations are like. The worst failure is that virtually all likelihood theorists (mentalistic *or* physicalistic)[14] depend on the ***rigidity principle,*** that is, that we perceive just that rigid 3-D object which fits the changing 2-D pattern of light in the eye or on the screen. This assumption can make it relatively easy to predict what surfaces or jointed structures (like people and animals) should be perceived, and why they are perceived; but the assumption is wrong. It has long been contradicted by laboratory demonstrations.[15] And it really does not apply to normally viewed moving pictures, in which changes in camera lens (close-up, long shot, and so on) and audience seating make it almost certain that ***only a nonrigid and deforming object*** could provide a geometrical fit to the moving image on the screen.[16] (Indeed, it seems likely to us that familiar animate bodily movements, and perhaps others as well, are perceived as opportunistic and therefore elastic fits to the motions of the extremities, rather than as rigid motions anchored at their joints, as has been claimed.[17])

There are many other problems that current accounts of mental representation leave unaddressed.[18] That is unfortunate, because as we see next the use of cuts in moving pictures poses a clear need for a theory that is specific about the nature of mental representations of events.

We take one last stab at a nonmentalistic account, and then a first step toward a mentalistic one.

Overlapping and Nonoverlapping Cuts

In most film and video, events and layouts are conveyed by both motion-based information and by discontinuous shots. Either could be used exclusively, but that is only rarely done. The Heider-Simmel film (like the scenes in Hitchcock's *Rope*) lies at one extreme, a single continuous shot, with no change in camera viewpoint. At the other extreme, Chris Marker's *La Jetée* (1964) consists of cuts between some 424 separate shots in 27 minutes, ***all but one of which contains no subject movement at all.*** (There are a few camera movements within still pictures.) It is an engrossing visual narrative, despite the absence of movement. What is important to us is that ***it is essentially a normal film in immediate memory, even as one watches it.*** And it is hard to see how one might hope to discuss an event or a layout as communicated in this way without assuming some contribution by the viewer, some mental representation.

Overlapping cuts (Fig. 17.3A) might conceivably challenge this assumption because they potentially specify physically how the camera or eye has moved relative to the scene as a whole. This physically specified factor might in fact contribute to how we combine our successive saccadic glances at the world,[19] but it fails to predict what viewers actually perceive in overlapping cuts. Both low-level and high-level processes can subvert the technique (when indeed it has any effects at all). Depending on spacing, the low-level mechanisms mentioned earlier can provide misleading apparent motion between noncorresponding objects in successive views: for example, in Fig. 17.3A, because each letter in the second shot (*ii*) is just rightward of noncorresponding letters in the first shot (*i*), a rightward jump is seen instead of the true leftward displacement of the letters. This is a common cause of bad cuts and affects the course of looking at filmed narrative.[20] In Figs. 17.3Bi and 17.3Bii, a high-

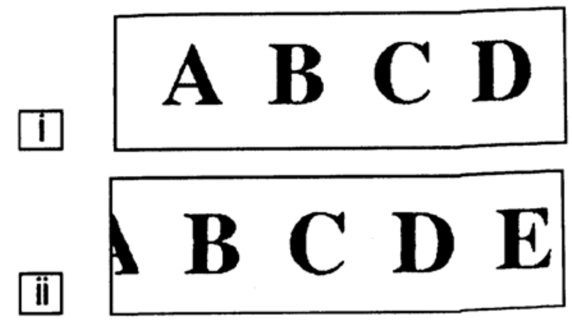

Fig. 17.3A. The overlap between shots *i* and *ii* specifies that *ii* is leftward of *i* in relation to eye or camera, but strong ***rightward*** motion is instead seen between the noncorresponding letters, thanks to low-level motion-detection mechanisms.

level cue as to what will come next, that is, the direction of the actors' gaze, overcomes the overlap between views (compare the long shot in Fig. 17.3C) and causes the second view (*ii*) to appear to lie rightward of the first (*i*), in the space beyond the screen.

For all we know at present, therefore, even overlapping cuts may work not because their overlap automatically "specifies" anything at all, but because overlap, like an actor's gaze, ***acts as a cue about what to expect.***

In any case, however, moving pictures also routinely use ***non***overlapping cuts. Such shots, by themselves, cannot convey anything about events or layouts beyond their boundaries. Admittedly, they can be individually remembered, to a degree. After being shown a rapid sequence of unrelated skills (in laboratory research), viewers can recollect information about some of the individual shots, and they show some signs of having visual expectations about what will come next. There is also evidence of a visual buffer that stores some small number of views;[21] indeed, we know that a great many briefly viewed pictures will be recognized on a second viewing as having been seen before. But recognition memory does not of itself provide the viewer with coherent events in rememberable sequence (or a rememberable place larger than the separate shots). Fig. 17.4Ai represents a stationary circular aperture on the screen (in the laboratory), through which the corners of a geometrical figure (a cross) are shown in sequence. Such sequences are not remembered. But if first shown a long shot as at Fig. 17.4Aii(1), the viewer can then test each successive view as to whether it fits the remembered cross in some regular order, and the sequence is then much better distinguished from other sequences.[22]

Without a mental structure in which to place the series of shots—their order as parts of an event, or as sample views of some spatial layout—the series is not rememberable. But it is rememberable given such a structure and the effort to apply it (that is, the attentional resources[23]). That argues that mental

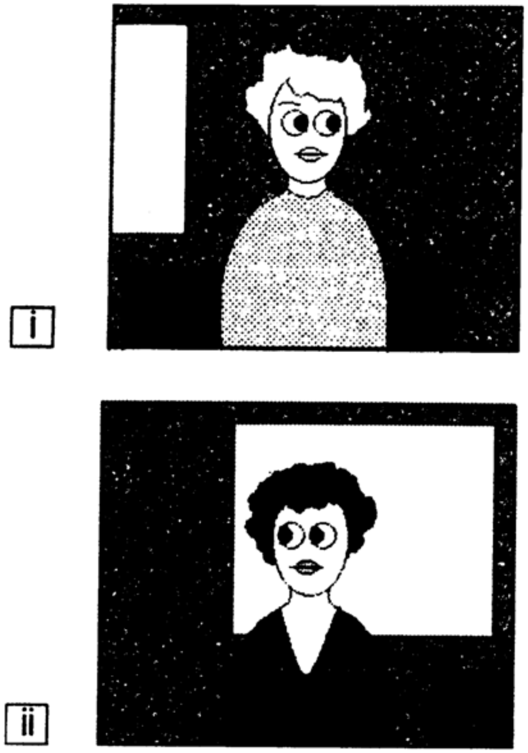

Fig. 17.3B. *i, ii.* Two successive shots, predominantly seen as a camera move from left to right.

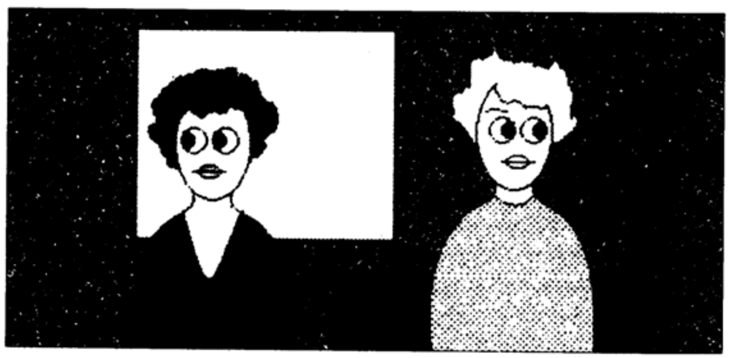

Fig. 17.3C. A more inclusive shot.

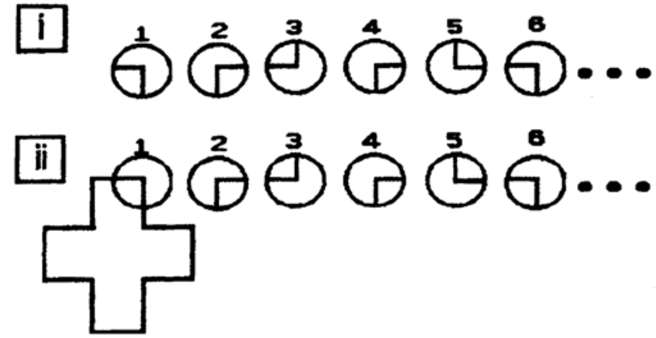

Fig. 17.4A. *i.* The first six shots in a sequence in which the corners of a cross are successively shown through a single stationary aperture. *ii.* The same sequence, with the entire cross shown in the first shot.

structure is involved in the process of event-perception itself, and we must therefore try to say more about what such structure is like.

Mental Representations and Story Structures

Something like this idea is found in many traditional cognitive theories, which take the testing of mental structures as central to the perceptual process.[24] To some filmmakers, good editing poses first a visual question and then a visual answer.[25] In experimental studies of film or video cutting, using short acted or animated scenes, both the visual question posed (as in Figs. 17.3Bi–iv) and the larger story structures within which the shots are presented significantly affect how the edited shots are comprehended.[26] But we must be more explicit about such mental structures if they are to be considered seriously.

Deciding what they *cannot* be like may help us to think of what they *may* be like. The mental structures fitted to our successive glimpses at the screen and at the world cannot have the characteristics of the world itself (despite assertions that they do).[27] Movements as we remember or anticipate them do not continue to run off in time, nor do remembered or anticipated layouts continue to extend in space. As we saw earlier, we start to lose extended time and space when the supporting input ceases. In any case, it certainly does not take some ninety minutes to review in our minds the average movie's representation.

We therefore simply cannot take either the moving picture or the events and space it represents as a faithful model for the film's mental representation. It is often hard to avoid doing just that, or to avoid making the opposite error by turning to the abstract story structure instead. There are plausible theories about written story structure, in the form of hierarchical analyses which account for much of readers' memories of the story's contents.[28] Bordwell has given narrative film a related approach, which explicitly deals with cuts, distinguishing the time covered by the story from the actual screen time after editing.[29] Despite some experimental applications suggested by this approach,[30] such analyses describe the structure of the narrative film itself, as we argue next, and not the viewer's mental representation while viewing the events on the screen.

Online Perception vs. Leisurely Analysis

Even with written stories, readers probably do not normally construct detailed representations while reading: After they have read that Mary stirred her cup, sensitive tests reveal no trace of a spoon in the readers' minds.[31] Nor is the entire story kept in mind and used on-line: subjects who were interrupted at reading regained their speed much faster after a brief review of the last few lines than they did after reviewing the preceding story structure.[32] While reading on-line, therefore, events that are distant in the narrative may go unconsulted, and unmentioned details are not filled in, so long as no inconsistency is encountered.[33]

When someone watches a film or tape unfold at its own pace, it seems even more likely that no filling in is done and no overall structure is consulted, so long as the new input is consistent with the immediately preceding or local context. And a well-made film, intended like most films in history for a single viewing, must therefore in general be *locally* comprehensible. In Fig. 17.4B, *i–iv* are sketches of shots 12–15 of *La Jetée,* stills of 1.27-2.0-, 1.0-, and 1.12-second durations, respectively. In watching the sequence, their relationship is evident without effort. Shot *i* poses a question, *ii* answers it (with an overlap that surely doesn't amount to physical specification); *iii* and *iv* are evidently looking at *ii,* following the principle in Fig. 17.3E. Using a continuous pan to connect the shots (as in Fig. 17.1) would have invited explanation in terms of relative motion detectors and invariants, but such explanations do not apply here (or to cutting in general). Giving the viewer appropriate acquaintance with the film prior to these shots would invite explanation in terms of story structure, but such reliance on overall structure would obscure the

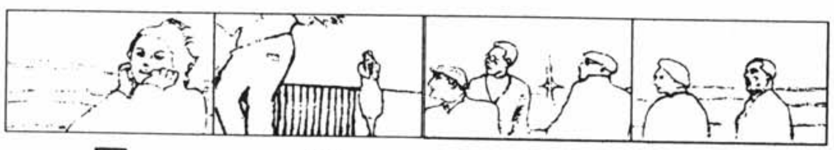

Fig. 17.4B. *i–iv.* Sketches of shots 12–15 in *La Jetée,* all stills, by Chris Marker.

purely local comprehensibility of the shot sequence. We do not know whether the local meaning depends on the actors' gaze as a cue (perhaps by directing the viewer's attention), or whether it reveals a relatively autonomous Gricean process of construal (for example, "why else are we being shown actors looking to the left?"). The final story structure will be *essentially the same* in any case. Only the films will differ, and the perceptual and cognitive processes by which their equivalent stories are achieved will differ too. That may be important for aesthetic reasons (pacing, affective cutting tone, attentional load, looking maintenance); it will be important for comprehension as well if the motions themselves are the film's subject, as in any filmed dance; and it is certainly important in any attempt to understand the perceptual processes at work in watching moving pictures. For many purposes, one might always patch such local analyses into the more conventional overall narrative analysis. But to do so one would first have to know about these local determinants, and whether they were in fact effective in the present instance.

As Thompson[34] has pointed out, there are many appropriate approaches from which to analyze film. To know what there is to comprehend in the film, to know what will arouse expectations early in the film that may later be fulfilled, to know what echoes and resonances are potentially effective, indeed to know what will be a misleading local reading of some stretch of the filmed narrative—these purposes and more require analyses in terms of overall narrative structure. Such analyses necessarily stand outside of the time flow of the film. To us, however, it is important to consider the film as it is experienced in the course of its first viewing (which for many a film will be its only viewing), thereby addressing what is most characteristic of the medium. Such an approach will distinguish one particular moving picture from another whose narrative might be summarized identically but which is very different in its moment-by-moment presentation. And such an approach scrutinizes the act of viewing itself, forcing us to attend to the rich mix of processes by which that act proceeds, and by which a mental representation of the narrated event is achieved, even though most of the steps along the way fade quickly in the course of the viewing.

In fact, considering how rapidly a filmed narrative can advance through brief and partial glimpses of previously unseen objects and events, perhaps at least three kinds of mental structure, each with different time scales and consequences, should be experimentally separated from story structure and from each other. First, there is the answer or confirmation to be obtained by the next glance. Second, there are the next expected features or landmarks, not immediately imminent, implied by the current action. Third, the viewer has a set of abstract readinesses, primed by previous events, for whatever may come afterward. Then, if a contradiction or some appeal by the film itself requires it, the viewer can consult and revise the story structure as far as it has developed. (Such appeals are very frequent. Why is that glass of milk so bright in *Suspense?* Why do we keep cutting away from the two combative adults to the rapt child in the *Third Man*? And why does the same child reappear among the crowd to stare accusingly at the camera/protagonist?) But this process will probably incur some cost in attentional resources, and escape the more passive viewer. The filmmaker, knowing the story structure, can only hope that the viewer can and will make the effort and draw the right conclusions.

It seems unlikely that, aside from the endorsement of some small repertory of local determinants (like that in Fig. 17.3B), such a detailed on-line analysis would be helpful to the narrative filmmaker. It seems likely that with all its limits, our storage capacities (as in speech) allow us to reconstrue earlier segments in the light of later information. But such analysis will certainly be helpful in extending present perceptual psychology, which is still largely confined to the study of the individual event, into the sequence of perceptual consequences with which it must deal in real-world behavior. It will probably be helpful in the programming of interactive (and virtual-reality) media. And it will certainly be helpful where the narrative itself is of a specific visual event, as in recording some dance performance, in which different cuts and camera angles may provide totally different visual (and audiovisual) experiences.

This brings us to the last point that emerges from our argument that the motions on the screen are not simply stored intact and subject to replaying. The story structure, as it is usually laid out by film scholar or cognitive psychologist in words and tree diagrams, also differs from what the viewer perceives and remembers, or can reconstruct, in that it is not visual. We need a notation system better fitted by visual displays than by words. Perhaps it should consist of brief high points or action features economically sampled from the flow of events; it will be relatively schematic, since details are not normally maintained unless needed; it will be mostly ego-centered, or camera-centered, with a definite viewpoint and 2-D composition as distinguished from an object-centered layout specified in 3-D coordinates; and the amount of information that it can store depends on how redundant the viewer finds the film. Remember that striking images, regardless of their place in the story hierarchy, are sure to be recognized when next viewed, and that can itself affect the story; a shot very similar to that of Fig. 17.4Bi reappears to great effect at the end of that film.

Comic strips and their predecessors[35] are a good approximation (although some frames serve only to hold dialogue); so is the filmmaker's storyboard, especially if it follows Eisenstein's shot scripting.[36] DeWied's ingenious recent version of Bordwell's analysis, combining storyboard, breakpoints, and tree structure, seems close to what is needed.[37]

Comic strips may be popular because they approximate the ways in which we think of the visual world. *La Jetée* may differ little in retrospect from what it would have been with full movements. Perhaps even a visual passage that exists only for its specific flow of movements, like a nonnarrative dance, remains with the viewer as an annotated "shot script." That representation may be hierarchical, in the limited sense that the knowledgeable viewer can, *if necessary,* reconstruct and insert additional boxes between the major break points; it may be that specific continuous motions are noted briefly where they are important (as in Figs. 17.1C, 17.1D); and it may even be that viewers can, when needed, reenvision those movements in real time.[38] Nevertheless, we take the mental representation that is approximated by this notation to be *in general* nonredundant and therefore static and discontinuous.

This is intended as first step at a description of the mental representation of visual events. The next step would seem to be close analysis and experimental research with moving pictures.

NOTES

1. Joseph Anderson and Barbara Anderson, "The Myth of Persistence of Vision Revisited," *Journal of Film and Video* 45 (Spring 1983): 3–12, and "Motion Perception in Motion Pictures," in *The Cinematic Apparatus,* ed. Teresa DeLauretis and Stephen Heath (New York: St. Martin's Press, 1980), pp. 76–95; David Bordwell, *Narration in the Fiction Film* (Madison: University of Wisconsin Press, 1985); Virginia Brooks, "Film, Perception and Cognitive Psychology," *Millennium Film Journal* 14/15 (Fall/Winter 1984–85): 105–26.

2. There seems to be no way to put this to rest. The latest version appears in J. Cantine, S. Howard, and B. Lewis, *Shot by Shot: A Practical Guide to Filmmaking* (Pittsburgh: Pittsburgh Filmmakers, 1993). For recent criticisms, see Anderson and Anderson, "Motion Perception"; Julian Hochberg and Virginia Brooks, "The Perception of Motion Pictures," in *Handbook of Perception,* vol. 10, ed. Edward C. Carterette and Morton P. Friedman (New York: Academic Press, 1978), pp. 259–304. For detailed history of where this incurable silliness started, see Olive Cook, *Movement in Two Dimensions* (London: Hutchinson, 1963) and Hugo Münsterberg, *The Film: A Psychological Study* (1916; reprint, New York: Dover, 1970).

3. There has been a recent explosion of work in this area. For discussion of short-range and longer-range motion mechanisms in relation to moving pictures, see Anderson and Anderson, "Myth of Persistence of Vision," and for a survey of research on the mechanisms and on the viability of the distinction, see Julian Hochberg and Virginia Brooks, "Perception of Moving Pictures Revised," in *Handbook of Perception,* ed. Edward C. Carterette and Morton Friedman, rev. ed. (in press).

4. Karl Duncker, "Über insuzerte Bewegung," *Psychologische Forschung* 12 (1929): 180–259. For a recent review, see Arien Mack, "Perceptual Aspects of Motion in the Frontal Plane," in *Handbook of Perception and Human Performance,* vol. 1, *Sensory Processes and Perception,* ed. Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas (New York: John Wiley, 1986), chapter 17, pp. 1–38.

5. The question of what visuomotor information the viewer takes into account when making pursuit movements, as when tracking a target in the world, and the complexities involved when a moving stationary viewer regards a stationary object against a moving background on a stationary motion picture screen, are more involved than we can treat here. See Hochberg and Brooks, "Perception of Moving Pictures Revised."

6. Gunnar Johansson, *Configurations in Event Perception* (Uppsala, Sweden: Almqvist and Wiksells, 1950). For a discussion of recent work on this phenomenon, see Hochberg and Brooks, "Perception of Moving Pictures Revised." For demonstrations that the effect is not simply the result of the eye's tracking *h,* leaving *v* as the only

motion on the retina, see Julian Hochberg and Peter Fallon, "Perceptual Analysis of Moving Patterns," *Science* 194 (1976): 1081–83.

7. With the stationary dot at either of the two corners which are not aligned with the diagonal, the vertical motion remains predominant, as reported by Julian Hochberg and Jeremy Beer, "Alternative Movement Organizations: Findings and Premises for Modeling (Abstract)," *Proceedings of the Psychonomic Society* (1990), p. 25; and as discussed in Hochberg and Brooks, "The Perception of Moving Pictures Revised."

8. Helmholtz's rule is compressed and rephrased from his *Treatise on Physiological Optic,* vol. 3, ed. and trans. J. P. C. Southall, from the 3rd German ed. (1909–11) (Rochester, N.Y.: The Optical Society of America, 1924–25), pp. 4–13. See Julian Hochberg, "Visual Perception," in *Stevens' Handbook of Experimental Psychology,* vol. 1, ed. R. Atkinson, R. Herrnstein, G. Lindzey, and D. Luce (New York: John Wiley, 1988), pp. 295–75. For work by current advocates of that position, see Richard Gregory, *The Intelligent Eye* (London: Weidenfeld and Nicolson, 1970); Irvin Rock, "The Logic of 'The Logic of Perception,'" (*Giornale italiano di psychologia* 20 (1994): 841–67; and Roger N. Shepard, "Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking, and Dreaming," *Psychological Review* 91 (1984): 417–47.

9. The strongest claim that perception is a direct response to stimulus information, not involving contribution from any mental representation, came from James J. Gibson, *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin, 1979). Without pursuing that claim, many visual scientists have pursued his goal of uncovering the rich, mathematically specifiable information about the world that is available to a moving observer (or one watching movies). Their recent work relevant to moving pictures is reviewed in Hochberg and Brooks, "Perception of Moving Pictures Revised." For somewhat opposing views of the information about movement into depth (a phenomenon relevant to the viewer's perception of dolly shots), see W. Warren and K. Kurtz, "The Role of Central and Peripheral Vision in Perceiving the Direction of Self-Motion," *Perception and Psychophysics* 51 (1992): 443–54; and James Cutting, *Perception with an Eye for Motion* (Cambridge: MIT Press, 1986). For a sophisticated study of how movement provides information about surface structure (for example, slopes, peaks, depressions), see J. S. Lappin and T. D. Wason, "The Perception of Geometrical Structure from Congruence," in *Pictorial Communication in Virtual and Real Environments,* ed. S. R. Ellis, M. K. Kaiser, and A. J. Grunwald (London: Taylor and Francis, 1991), pp. 425–48.

10. Jeremy M. A. Beer, "Perceiving Scene Layout through an Aperture during Visually Simulated Self-Motion," *Journal of Experimental Psychology: Human Perception and Performance* (in press).

11. Lynn A. Cooper, "Mental Models of the Structure of Three-Dimensional Objects," in *Object Perception: Structure and Process,* ed. B. Shepp and S. Ballestreros (Hillsdale, N.J.: Lawrence Erlbaum, 1989), pp. 91–119.

12. Fritz Heider and Marianne Simmel, "An Experimental Study of Apparent Behavior," *American Journal of Psychology* 57 (1944): 243–59.

13. D. Newtson and G. Enqvist, "The Perceptual Organization of Ongoing Behavior," *Journal of Experimental and Social Psychology* 12 (1976): 436–50; D. Newtson, J. Hairfield, J. Bloomingdale, and S. Cutino, "The Structure of Action and Interaction," *Social Cognition* 5 (1987): 121–237.

14. See Irwin Rock, "The Logic of 'The Logic of Perception'" and Shepard, "Ecological Constraints on Internal Representation" for mentalistic theories, and see Gibson, *Ecological Approach to Visual Perception;* G. Johansson, "Visual Space Perception through Motion," in *Tutorials on Motion Perception,* ed. A. H. Wertheim, W. A. Wagnaar, and H. W. Liebowitz (New York: Plenum, 1982); and Stephen Ullman, *The Interpretation of Visual Motion* (Cambridge: MIT Press, 1979) for physicalistic theories.

15. Strong evidence that rigidity is not the fundamental constraint that allows us to recover three-dimensional structure from changing images, to the exclusion of static pictorial depth information, is reported by M. L. Braunstein and G. J. Andersen, "Testing the Rigidity Assumption: A Reply to Ullman," *Perception* 15 (1986): 641–44; Julian Hochberg, "Machines Should Not See as People Do, But Must Know How People See," *Computer Vision, Graphics, and Image Processing* 37 (1987): 221–37; B. J. Schwartz and G. Sperling, "Non-Rigid 3D Precepts from 2D Representations of Rigid Objects," *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 24 (1983): 239 (abstract). The strategic value that the rigidity constraint offered direct theories of perception was that it made static pictorial depth cues unnecessary to any account of the perception of moving pictures, or of the perception of the world by a moving observer. Once we know that rigidity does not make the depth cues unnecessary, it becomes very hard to dispense with mental representations in one's theoretical account.

16. See Julian Hochberg and Virginia Brooks, "Perception of Still and Moving Pictures," in *International Encyclopedia of Communications,* ed. Erik Barnouw (New York: Oxford University Press, 1989); Julian Hochberg, "Representation of Motion and Space in Video and Cinematic Displays," in *Handbook of Perception and Human Performance,* vol. 1, ed. Kenneth R. Boff, James P. Thomas, and Lloyd Kaufman (New York: John Wiley, 1986), pp. 1–64.

17. For their explanation of the well-known film of actors made visible only by small lights at their joints, see G. Johansson and G. Jansson, "Perceived Rotary Motion from Changes in a Straight Line," *Perception and Psychophysics* 4 (1986): 165–70.

18. See Julian Hochberg, "Perceptual Theory and Visual Cognition," in *Cognitive Approaches to Human Perception,* ed. S. Ballesteros (Hillsdale, N.J.: Lawrence Erlbaum, 1994), pp. 269–89.

19. This was suggested by J. J. Gibson, *Perception of the Visual World* (Boston: Houghton Mifflin, 1950). It is the only explicit purely visual explanation offered for how glances combine, and it has some recent experimental support. (See D. F. Irwin, J. L. Zacks, and J. H. Brown, "Visual Memory and the Perception of a Stable Visual Environment," *Perception and Psychophysics* 47 [1990]: 35–46.) But it has also allowed Gibson and his followers to argue that the entire optic array, which remains invariant under the changing glance of eye or camera, is the effective stimulus on which perception is based—an argument that has no experimental support and much to oppose it. (See Hochberg, "Visual Perception.")

20. Julian Hochberg and Virginia Brooks, "Film Cutting and Visual Momentum," in *Eye Movements and the Higher Psychological Functions,* ed. J. W. Senders, D. F. Fisher, and R. A. Monty (Hillsdale, N.J.: Lawrence Erlbaum, 1978), pp. 293–313; G. D'Ydewalle and M. Vanderbeeken, "Perceptual and Cognitive Processing of Editing Rules in Film," in *From Eye to Mind: Information Acquisition in Perception, Search, and Reading,* ed. R. Groner, G. d'Ydewalle, and R. Parham (Amsterdam: North Holland, 1990), pp. 129–39.

21. The experimental study of montages of unrelated stills deserves more attention in film studies than it has received. For its start, see M. Potter and E. Levy, "Recognition Memory for a Rapid Sequence of Pictures," *Journal of Experimental Psychology* 81

(1969): 10–15. For a recent review and analysis of what the data imply as to mental representations and the normal integration of our successive glances, see H. Intraub, R. Bender, and J. Mangels, "Looking at Pictures But Remembering Scenes," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (1992): 180–91; and H. Intraub, "Contextual Factors in Scene Perception," in *The Role of Eye Movements in Perceptual Processes,* ed. E. Chekaluk and K. R. Llewellyn (Amsterdam: North Holland, 1992), pp. 47–72, respectively.

22. As described in Hochberg, "Representation of Motion and Space in Video and Cinematic Displays," pp. 58–60.

23. Although we know of no experimental research to this point, it is clearly a demanding task, requiring both intention and concentration.

24. Such structure testing underlies the general formulations offered by psychologists and philosophers of the past and present centuries, starting with John Stuart Mill and von Helmholtz.

25. Karel Reisz and Gavin Millar, *The Technique of Film Editing,* 2d ed. (New York: Focal Press, 1968).

26. R. N. Kraft, "The Influence of Camera Angle on Comprehension and Retention of Pictorial Events," *Memory and Cognition* 25 (1987): 291–307; P. S. Cowen, "Manipulating Montage: Effects on Film Comprehension, Recall, Person Perception, and Aesthetic Responses," *Empirical Studies on the Arts* 6 (1988): 97–115; R. N. Kraft, "Light and Mind: Understanding the Structure of Film," in *Cognition and the Symbolic Processes: Applied and Ecological Perspectives,* ed. R. R. Hoffman and D. S. Palermo (Hillsdale, N.J.: Lawrence Erlbaum, 1991), pp. 351–70; d'Ydewalle and Vandebeeken, "Perceptual and Cognitive Processing."

27. Rock, "The Logic of 'The Logic of Perception'"; Shepard, "Ecological Constraints."

28. G. Bower, J. Black, and T. Turner, "Scripts in Memory for Text," *Cognitive Psychology* 11 (1979): 177–220; J. M. Mandler, "A Code in the Node: The Use of a Story Schema in Retrieval," *Discourse Processes* 2 (1978): 14–35; J. M. Mandler and N. S. Johnson, "Remembrance of Things Parsed: Story Structure and Recall," *Cognitive Psychology* 9 (1977): 111–51.

29. Bordwell, *Narration in the Fiction Film.* When Bordwell's distinctions between the various scales of narrative time in moving pictures are placed against the limited resources of working memory, something like the mental shot script we discuss below seems an unavoidable attribute of on-line mental representation.

30. M. A. de Wied, "The Role of Time Structures in the Experience of Film Suspense and Duration: A Study on the Effects of Anticipation Time upon Suspense and Temporal Variations on Duration Experience and Suspense" (Ph.D. diss., University of Amsterdam, 1991).

31. B. A. Dosher and A. T. Corbett, "Instrument Inferences and Verb Schemata," *Memory and Cognition* 10 (1982): 531–39.

32. M. Glanzer, B. Fischer, and D. Dorfman, "Short-Term Storage in Reading," *Journal of Verbal Learning and Verbal Behavior* 23 (1984): 467–86.

33. G. McKoon and R. Ratcliff, "Inference during Reading," *Psychological Review* 99 (1992): 440–66.

34. Kristin Thompson, *Breaking the Glass Armor: Neoformalist Film Analysis* (Princeton: Princeton University Press, 1988).

35. E. H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Princeton: Princeton University Press, 1961).

36. Vladimir Nizhny, *Lessons with Eisentein,* trans. and ed. Ivor Montagu and Jay Leyda (New York: Hill and Wang, 1962), pp. 62–92.

37. De Wied, "Role of Time Structures."

38. Beer, "Perceiving Scene Layout"; Cooper, "Mental Models"; Lynn A. Cooper, "Demonstration of a Mental Analog of an External Rotation," *Perception and Psychophysics* 19 (1976): 296–302.