

to appear in  
Joseph D. Anderson and Barbara Fisher Anderson (eds.),  
*Moving image theory: Ecological considerations*  
in press.

## **Perceiving Scenes in Film and in the World**

James E. Cutting  
Cornell University

### **Abstract**

Watching a film is different than observing the real world. In particular, scenes in film are framed and set off from a larger context, they are divided up into shots that are composed from different points of view, separated by instantaneous cuts, and with the camera performing other feats impossible for the unaided eye, such as zooming in. Real life has none of this. How is it that we come to accept the wholeness and integrity of film with multiple shots and cuts? Given that we never evolved to see this structure, it is curious that it works so well. The reasons for film's success, I claim, stem from our biological endowment, how it constrains and does not constrain our cognitive and perceptual systems in dealing with space and time. Directors and cinematographers exploit this in what is called Hollywood style.

The real world is spatially and temporally continuous; film is not. We evolved in a continuous world and, regardless of how much we may enjoy them, we emphatically did not evolve to watch movies. Instead movies evolved, at least in part, to match our cognitive and perceptual dispositions. The result is a curious melange of short shots with instantaneous camera jumps between them, something not at all like the rest of the world around us. Why and how do we accept this? Part of the answer, I claim, is that we do not necessarily perceive the world according to its physical structure. For example, although we evolved in a Euclidean world, our perceptions of space around us are generally not Euclidean, and generally don't need to be (see Cutting & Vishton, 1995; and Cutting, 1997, for more discussion). In addition, although we evolved in a temporally continuous world, our perception of time is not tightly bound to any temporal meter. Thus, there is a considerable plasticity to our perceptual world; its just happens that the world is mostly rigid and evenly flowing.

Part of the success of film can be attributed to the goals of what is sometimes called *Hollywood style* (see Bordwell, Staiger, & Thompson, 1985).<sup>1</sup> Without endorsing any political or social aspects of this genre, one finds that Hollywood style has a main goal that is almost purely cognitive and perceptual—to subordinate all aspects of the presentation of the story to the narrative (e.g. Messaris, 1994; Reisz & Miller, 1968). This means that, generally speaking, all manipulations of the camera, lighting, editing, and sets should be transparent, unnoticed by the filmgoer. To go unnoticed these techniques must mesh with the human visual system. Finally, to understand why film works so well is to understand much about how we perceive the real world; and to understand how we perceive the world tells us much about how we understand film. This is, I claim, the fundamental tenet of an ecological approach to cinematic theory.

This chapter is about our perception of space (or, better, layout) in the world and in film; and then of how space and time can be cut up to make a film scene. But first let me establish some terminology. Film is made up *shots*, each consisting of the continuous run of a camera. For 75 years the maximum standard shot length for 35-mm film has been 10 minutes—1200 feet of film, or the running time of one standard reel—although few shots are ever that long. In the production of a typical Hollywood film, shots tend to be much longer in the initial photography, and then in the editing process each shot is trimmed to a few seconds in length for the final film. The shot is then juxtaposed, without transition, with another shot taken from another point of view. This juxtaposition is called a *cut*. A *scene* usually takes place in a single location. Typical scenes are made up of many shots and cuts; and most films, of course, are made up of many scenes. We don't usually speak of real life as being made up of scenes, but it does no real injustice to speak this way. We walk through life. When our environs are roughly the same, such as when strolling outside across a town square, we could call this a scene; when they change, perhaps when we then enter a building, we could call this a break between scenes, going on to the next. Episodic memory, a central concern of cognitive science, is essentially the memory of scenes from our life.

In film, every shot shows an environment of some kind. This environment has a physical arrangement, or a layout, of objects and people. The projection of this layout is

unique to a particular camera position, but as viewers we pay little attention to this projection. Instead, we focus on the “world behind the screen.” We also view the real world at any given time from a particular position, and also generally ignore its particular projection to our eyes, focusing instead on the general 3D layout of the environment. Cinematographers and film directors, of course, pay considerable attention to camera position, crafting the composition of the image. In particular, they manipulate the information available to portray the layout of the scene as they deem best. What are they manipulating? Consider an answer in terms of contemporary and traditional research in the visual sciences.

### **How Layout (Depth) is Revealed through Different Information**

To begin, it will be useful to separate nine of the different sources of information (traditionally called “depth cues”) available to an observer in the real world, and then apply these sources to film. Few if any of these sources, by themselves, imply a metric space (measured in ratios and absolute distances). Although, in consort all can contribute to a near-Euclidean representation of space relatively near us under ideal conditions, there is enough leeway for a seasoned cinematographer and film director to carve out of them more or less what he or she wants us to see. Consider each in turn, applied to the world and then to film

**1. Occlusion** occurs when one object partly hides another from view. Cup one hand in the other, and the hand closer to your eyes partially occludes the farther. As an artistic means of conveying depth information partial occlusion has been found in art since paleolithic times, where is it often used alone, with no other information to convey depth. Thus, one can make a reasonable claim that occlusion was the first source of information discovered and used to depict spatial relations in depth. And of course it is found in the earliest photographs and films as well. However, occlusion is never more than ordinal information—one can only judge that one object is in front of another, but not by how much. Thus, the kind of space that can be built up from occlusion information alone is an affine space—one that can squash, stretch, and shear. Camera position and the layout of clutter in a scene will dictate to the observer (and camera) which objects occlude or partly occlude other others. If only occlusion occurs within a shot, a perceiver will not be able to know exactly where two objects are. It gives great power to the cinematographer.

Occlusion is unavoidable in film, so much so we often take it for granted. We should not. It is used very effectively, for example, in a temporal-lapse sequence in Roger Mitchell's 1999 film *Notting Hill*. Between flirtatious episodes with movie star Anna Scott (Julia Roberts), bookseller Will Thacker (Hugh Grant) walks through the market in London's Notting Hill, being occluded by arcades, stands, and people. The sequence appears continuous and the camera follows Thacker with a long tracking movement, most of it with the camera's line of sight at 90° to its motion. Seasons change through a full year during the stroll and track, juxtaposing two types of time—that measured in seconds with that measured in months. Given that the camera follows Thacker, our attention remains on him even when he is out of sight. Among other things, this demonstrates that objects at different depths but at the same retinal location can be attended to separately, an idea that has received much laboratory focus (see Atchley,

Kramer, Andersen, & Theeuwes, 1997, for a review; see also Neisser & Becklen, 1975). Despite appearances, this *Notting Hill* sequence is not a continuous shot. Manipulating the viewer's attention, the editor uses occlusion to hide a cut in the shot transitions, which is necessary for the circular movement of the camera in the second part of the sequence, a fine example of following Hollywood style. A similar solution to a technical problem is used in Hitchcock's 1948 film *Rope*, about which more will be said later.

**2. Height in the visual field** concerns object positions in the field of view, or in the frame. Objects occupying higher positions are generally farther away. This information typically measures relations among the bases of objects in a three-dimensional environment as projected to the eye or camera. Like occlusion, height in the visual field offers only ordinal information, and like occlusion it has been used in pictorial representations since near the beginning of art. Moreover, with photography and film, camera height is often manipulated for specific effect. A high, downward-tilting camera reveals greater differences among the bases of ground plane objects measured in the picture plane, giving more articulated information. A low and level camera, on the other hand, diminishes the availability of this information, forcing us to compare object juxtapositions without height information. A high and level camera yields the same kind of distance information as lower one but farther out into space, giving a grander view.

Relations among objects in terms of height in the frame reciprocally specify the height of the camera and the camera angle with respect to the ground plane. The height of the camera and its angle, in turn, place the perceiver in a subjective position—high often indicating dominance (as with adults looking down at children), and low a more submissive role (as with children to adults; e.g., Messaris, 1994). The first half of Robert Wise's 1965 film *The Sound of Music* is largely about the Von Trapp children. It is shot mostly from an eye height slightly less than an adult. The second half of the film, however, is largely about the romance between Maria, the governess, and Captain Von Trapp. It is shot mostly from the eye height of an adult. Indeed, viewers are supposed to identify with the children in the first half of the film (Maria is also winning us over as governess) and with the adults in the second. The point here is that the relations among objects, particularly as revealed by height in the picture plane, tell us where our eye is—and thus help tell us whether we, the film audience, are "children" or "adults." We don't notice this watching the film; it is a part of Hollywood style.

In his 1957 film *Twelve Angry Men*, Sidney Lumet used systematic differences in camera height across the course of the film, manipulating information (among other things) about height in the visual field (Lumet, 1995). Unlike Wise's film, dominance and identification are not primary factors here; manipulation of space is. Roughly the first third of the film was shot at a standing eye height which, since most of the actors are sitting at the jurors' deliberation table, gives ample information about the locations of objects on the table and positions of individuals around it. The second third of the film was shot generally at a sitting eye height. This foreshortens the table and makes less clear where things and people are, but we already know this because of the first part of the film, and the lower camera height draws us into the deliberation around the table. And the final third of the film was shot just below sitting eye height, removing the plane of the table almost completely. This deletes the space in front of the individual

jurors, isolating them from in their deliberations from their locations at the table and thus from each other. But again, we don't notice this manipulation.

**3 & 4. Relative size and relative density** concern how big objects are, and how many there are, as seen by the eye. Pebbles are large and not numerous seen held in the hand, but they are smaller and more numerous seen on a rocky beach. More technically, relative size is a measure of the angular extent of the retinal (or image) projection of two or more similar objects or textures. It has been used in some rough sense since at least early Greek, if not Egyptian and Persian, art. Unlike occlusion and height in the visual field, relative size has the potential of yielding ratio information. That is, for example, if one sees two similar objects, one of which subtends one half the visual angle of the other, the former will be twice as far away. Technically, relative density concerns the projected number of similar objects or textures per solid visual angle, and is what Gibson (1950) meant by term texture gradient. It works inversely to relative size, and is considerably weaker in its perceptual potency (Cutting & Millard, 1984; Cutting & Vishton, 1995). Relative density is a relative latecomer to art; its effects were first seen in the local (not fully coherent) perspective piazzas of the 14<sup>th</sup> century. Its lateness to the armamentarium of depiction is due to the fact that only with the invention and use of linear perspective in Renaissance art are these first four sources of information—occlusion, height, size, and density—coupled in a rigorous fashion, and the technology of depicting density differences is the hardest to carry out. Unlike relative size but like the first two sources, relative density provides only ordinal information about depth. Computer graphics allows independent manipulation of relative density and relative size, but with a camera in the real world, the two are yoked: As size of texture elements doubles, their density decreases by half.

In photography relative size and density are manipulated through the use of lenses (e.g. Swedlund, 1981). Perhaps the most familiar example of issues concerning relative size occurs in portrait photography. Here the photographer typically stands back from the subject and uses a long lens. For 35-mm film, the standard lens has a focal length of 50 mm, and a lens with a focal length greater than about 100 mm is considered a long lens, also called a telephoto lens. With a short focal length lens on the camera, the camera must be placed close to the person being photographed, with the result that the difference between camera-to-nose distance and camera-to-ear distance is great, and the person's nose appears large. With a long lens on the camera, the camera can be placed farther away from the person being photographed. With the camera farther away the difference between camera-to-nose and camera-to-ear distances becomes negligible, so the person's features appear close to their actual sizes. This is also one reason why most shot-reverse-shot sequences in cinematic dialogs are taken with relatively long lenses. They make the actors look better. More on dialogs later.

Manipulations of relative size through lenses have other important effects, dilating and compressing space with short and long lenses, respectively. One memorable scene near the end of Mike Nichols' 1967 film *The Graduate* has Benjamin Braddock (Dustin Hoffman) running down a sidewalk, his car having broken down, trying to stop the wedding of the young woman he loves. He runs for more than 10 seconds directly towards the camera (into a very long lens), with the appearance of getting nowhere. This getting-nowhere effect—enhancing the anxiety of the viewer—is

conveyed by the fact that the long lens compresses depth. This compression results from decreased differences in relative size (and density) in the sections of sidewalk and the surrounding trees and bushes, and also keeping Braddock from growing much in size as he strains to get to the church.

Such spatial compression and dilation effects find themselves useful in many situations. Again in *Twelve Angry Men*, Lumet shot the first third of his film with relatively short lenses, dilating depth, and conveying a wide-angle spaciousness of the deliberation room. He then shifted to more standard lenses in the next third; and a long lenses in the final third, narrowing the field of view and compressing the space around the jurors as the debate progressed, creating more tension. Combined with the progressively lower camera angles, by the end the ceiling is revealed to be pressing in on the jurors as well. But again, all of this is unnoticed; we follow the narrative and the lens effects support the narrative.

Perhaps the most striking spatial transformation is attributable to Alfred Hitchcock (e.g. Truffaut, 1983) in a sequence that gives eponymic visual force to 1957 film *Vertigo*. Hitchcock wished to simulate Scottie Fergusson's (James Stewart's) fear of heights during his views down a belltower's stairs. The effect is done in a subjective shot (one following an objective shot of Ferguson looking down the stairwell, and called point-of-view editing) by combining a dolly in with a zoom out.<sup>2</sup> This procedure keeps the near steps the same size, but dilates the space changing the apparent depth by changes in the relative size of farther objects but not nearer ones. The scene has a stomach-churning plastic and deforming character. The bottom of the stairwell rushes away from the viewer, getting deeper and more dangerous.

It should be noted that the effectiveness of this dolly/zoom depends on the viewer having *some* near-metric information about depth. If the visual system's ability to deal with depth were completely plastic (affine) such effects would not be noticed at all! Thus, there is a sense in which we notice the effect. This would appear to conflict with the idea of Hollywood style, but it does not. The effect is the key element of the narrative. Ferguson has vertigo and we, personally, can see that it is an awful and debilitating thing.

**5. Aerial perspective** refers to the effects of fog, mist, and haze. These create an indistinctness of objects with distance determined by moisture or pollutants in the atmosphere. Its perceptual effect is a decrease in contrast of the object against the background with increasing distance, converging to the color of the atmosphere. Aerial perspective was systematically discussed and understood by Leonardo (Richter, 1883), and has been used photographically and cinematographically since their beginnings. Like many other sources, it is ordinal information. Objects are dimmer and less distinct when farther away, but as viewers we don't know by how much because we really cannot accurately assess the density of the atmosphere.

In photography and film, aerial perspective (particularly as fog) can also be manipulated with lenses. Long lenses bring more of the atmosphere into play between objects in focus and in the field of view. The final scenes of Michael Curtiz's 1941 film *Casablanca* use lenses and fog quite effectively. Wondering who, if anybody, will escape

the Nazis from this North African city, the audience sees—because of the fog—the airplane at a barely attainable distance behind Rick Blaine (Humphrey Bogart), Ilsa Lund (Ingrid Bergman), and Victor Laszlo (Paul Henreid). This heightens the viewer's anxiety about possible departure.

But perhaps the clearest example of this effect is not with fog but with rain. The bulk of televised baseball games is typically shot with long lenses from behind the catcher or from center field. These show the pitcher, batter, and catcher occupying what seems to be the same space, nearly on top of one another. This, of course, is the effect of relative size compressing depth discussed earlier. But on nights with a smattering of rain, not enough to stop the game, the images shot with long lenses make the scene look like a veritable downpour. This is *not* Hollywood style, since one wonders why the umpires do not stop the game. Yet the downpour is a false impression—more raindrops in depth are compressed into the field of view than are experienced by ballplayers.

**6. Accommodation** occurs with the change in the shape of the lens of the eye, allowing it to focus on objects near or far while keeping the retinal image sharp. Objects at other distances are blurred. The camera analog to accommodation occurs with dynamic manipulation of focal depth, which can place one object in focus and another out. This information tells the viewer only that the objects are at different depths. By itself, however, it does not even tell depth order. Interestingly, blur first appeared in art about the same time with Impressionism and with late 19<sup>th</sup> century photography (Scharf, 1968).

Manipulation of clear and blurred regions of an image is also a powerful tool for the cinematographer. It is used to control points of interest in a scene, where he or she wants the viewer to look. This is done effectively, for example, in *The Graduate* when looking over Benjamin Braddock's shoulder and first focussing on Elaine Robinson (Katherine Ross) in her bedroom then on Mrs. Robinson (Anne Bancroft) in the more-distant hallway. Only one is in focus at a time. They are thus revealed at different distances, and the narrative's sequence of outrage passes from one to the other.

**7 & 8. Convergence and binocular disparity** are two-eyed phenomena. Convergence is measured as the angle between foveal axes of the two eyes. When large, the two eyes are canted inward to focus near the nose; when approaching 0° the two eyes are aligned to focus beyond 10 m (which is, interestingly enough, functionally the same as the horizon). Convergence can be registered and used at close range, but not beyond about 2 m. Given that photographic and cinematic images are flat it is uninformative, and given that all of film and much of television is watched from distances greater than 2 m this source of information is irrelevant. Binocular disparities are the differences in relative position of sets of objects as projected on the retinas of the two eyes. When disparities are sufficiently small they yield stereopsis, or the impression of solid space. When disparities are greater than stereopsis will allow, they yield diplopia—or double vision—which is also informative about relative depth. Stereo is also extremely malleable, and just one day of monocular vision can render one temporarily stereoblind (Wallach & Karsh, 1963).

Convergence has never had artistic use and it is remarkable that stereo has never played an important role in photography or film, except as a type of parlor teaser. Despite all predictions at the time and before (see Eisenstein, 1948/1970), few pictures following the 1953 film *House of Wax* by André de Toth, have been made in 3D. Some theorists suggest that the reason has to do with the relative gimmickry of stereo and the necessity of wearing glasses (e.g. Kubovy, 1986). Without denying this factor, I think stereo films fail as an important medium because stereo in the real world enhances noticeable depth differences only nearest to the viewer and, as I will discuss later, this is not a region of space that is important to most filmmakers. Interestingly, Hitchcock's 1954 film *Dial M for Murder* was shot in 3D but is rarely seen in this format; Hitchcock had particular interest in low camera angles in this film and 3D worked well to reveal depth differences in near space (Truffaut, 1983, p. 210). However, perhaps unsatisfied with its effects, Hitchcock never used 3D again.

Convergence and disparities are linked in human vision, but having two eyes can often get in the way of seeing depth in pictures. At the turn of the 20<sup>th</sup> century Karl Zeiss, inventor of cameras and of the planetarium, hoped to gain (another) fortune by selling a device that neutralized both sources for visitors to art museums. Called a synopter, this apparatus contains a series of fully silvered and half-silvered mirrors at 45° angles that superimpose the lines of sight of the two eyes, nullifying disparities and convergence. Reports suggest these devices greatly enhance the visual depth seen in photographs and in paintings (Koenderink, van Doorn, & Kappers, 1994), often even more than stereoscopic displays. The reason for enhanced depth compared to unencumbered two-eyed viewing seems relatively straightforward. This device cancels certain information about flatness (uniformly graded disparities, and vergence on a nearby object) in the scene viewed. It also can remove from view the frame and other context surrounding the picture.

The reason the synopter produces an effect of depth *better* than stereopsis is a bit more complicated. First, with typical stereo material there is a *coulisse* effect (objects can appear relatively flat with startling spatial gaps in depth between them). This effect is due to the fact that the two cameras used to take the stereo images are usually considerably more than 6 cm apart, the distance between our two eyes. Stereo cameras wider apart than our eyes will tend to minify a scene (they effectively enlarge our "head"), making objects appear proportionately smaller and flatter than they are. Indeed, early parlor stereograms were of European cities, taken with cameras as much as a half-meter apart (8 times normal) or more. This renders impressions of the cities very much less grand (1/8 the size), even toylike. Second, zero disparity, which is what the synopter achieves, does not actually take away depth information. Instead, it specifies infinite depth, or at least a depth beyond about 30 m. This would probably be pooled with other sources and enhance the overall depth effect. Most simply, however, moviegoers can achieve a nearly synoptic effect by sitting more than 10 m from the screen.

**9. Motion perspective** refers to the field of relative motions of objects rigidly attached to a ground plane around a moving observer or camera. It specifically does not refer to the motion of a given object, which was the major early accomplishment in the ontogeny of film.<sup>3</sup> Motion perspective occurs best during a dolly (or tracking shot),

where near objects and textures move faster than far ones, and their velocity is inversely proportional to their distance from the camera. Thus, objects twice as far move exactly half as fast, so long as the camera does not pan. The first uses of motion perspective in film were seen at the end of the 19<sup>th</sup> century (e.g. Toulet, 1988), where cameras were mounted on trolleys and trains, and their effects presented to appreciative audiences.

Motion perspective is particularly good at generating the impression of self-movement, but it needs to be distinguished from another camera manipulation. In early and later cinema outside the studio dollies entailed putting a camera in a moving vehicle or, more expensively, the laying down of a track on which the camera rolled. For example, the filming of the background in the chase sequence through the Ewok (redwood) forest near the end of George Lucas's 1983 film *Return of the Jedi* used a track and a dolly camera. He also used frame-by-frame photography to enhance the speed, and then hand blurring of the periphery of each image to avoid motion aliasing artifacts. Today steadycams (cameras using inertia to avoid the bounciness of hand-held techniques) make the motion-perspective effect easier to attain outside the studio.

The information about motion perspective attained from a dolly should be distinguished from the patterns seen in a zoom. Zooming in, as suggested above, is the continuous adjustment of a variable lens from relatively a short to a relatively long length (the range of 38 to 115 mm is common in a 35-mm camera). The optical differences between the two are interesting, but in short sequences is generally unnoticed by a film viewer (Hochberg, 1978). Zooming in simply enlarges the focal object, allowing all texture to rush by with equal speed as a function of its image distance from the center of the focal object. No occlusions and disocclusions occur in a zoom. Motion perspective, on the other hand, creates occlusions and disocclusions of far objects by near ones, and the objects and textures rush by at a speed proportional to their physical distance from the camera and their angle from the path of the camera.

Although viewers may not be, filmmakers are quite sensitive to the differences between a dolly and zoom. Dollies are used to indicate observer motion; zooms are typically used for increased attention. Interestingly the phenomenon of attention as studied within experimental psychology generally supports this idea. Attention is a phenomenon of increased interest on an object, typically in the center of the field of view, coupled with an increased rejection of information in the periphery. Indeed, theories of attention occasionally talk, metaphorically, of zooming in on objects during periods of interest (see Palmer, 1999, for a review of the phenomena of attention).

**Shadows and lighting** are often added to lists of information contributing to the perception of depth. However, I believe shadows are used almost exclusively for articulating the shapes of objects, not about object relations in depth around the perceiver. The reason is straightforward—changes in shadows rarely change depth or one's perception of the shape of an object, whereas changes in relative size, height in the visual field, binocular disparities, and the like almost always do. This is not to underplay the importance of shadows in real life or in cinema. Artistically it is crucial to play with the identity of objects and individuals in film, and this is often done best through variations in lighting. Consider two particularly striking examples of lighting

effects. One occurs in Hitchcock's 1938 film *The Lady Vanishes* where a handwritten message in the condensation on the interior of a train window is invisible in daylight, but appears when the train is in a tunnel—a key bit of evidence on which turns possible hallucination into intrigue. A second occurs throughout Godfrey Reggio's 1982 film *Koyaanisquatsi*, a film without dialog or standard plot. Time-lapsed photography is used throughout with the camera often remaining in position throughout a full day, recording the passing of events under the change of light.

### Phenomenal Spaces in the Real World and in Cinema

On the basis of the differential relative potency of the various sources of information listed above, I have found it convenient to divide egocentric space into three regions—vista space (that beyond about 30 m) for a pedestrian, action space (from 30 m inward to about 1.5 m), and personal space (closer than about 1.5 m; see Cutting & Vishton, 1995; Cutting 1997). In vista space the only effective sources of information are the traditional "pictorial cues"—occlusion, height in the visual field, relative size, relative density, and aerial perspective—all of which are yoked within the technique of linear perspective mastered by Renaissance artists, and yoked in camera use as well. Motion perspective for the pedestrian is not particularly effective beyond 30 m, particularly when looking in or near the direction of motion. Similarly, stereo is also not very effective. Vista space can be strikingly portrayed in large *trompe l'oeil* paintings and in cinema, particularly in wide-screen format. But the typically narrative content of vista space in film is nil. Vista is only backdrop, and older Hollywood movies succeeded well by simply painting vistas on walls and on movable sets.

Action space is circular, on the ground plane around us, and generally closer than about 30 m but beyond arm's reach. We move quickly within this space, talk within it, and toss things to a friend. More simply, we act within this space. In everyday life, all but the most intimate conversations occur within this space. In the real world this space appears to be served by a different collection of information sources: three of the five linear perspective sources (relative density and aerial perspective are usually too weak compared to the others) plus binocular disparity and motion perspective. For film we can omit disparities. Most emphatically, action space is the space of films. Film content almost always takes place between 2 and 30 m of the camera. As viewers, we like it this way.

The near boundary of action space for the pedestrian is delimited by the emergence of height in the visual field as a strong information source, which also serves to limit this space to the ground plane. Viewing objects from above or below about 1.5 to 2 m tends to make perception of their layout less certain by weakening the effect of *familiar size*, a phenomenon by which we can scale the size of surrounding objects by what we know to be the size of a particular object. Many wide-angled paintings and engravings from the 18<sup>th</sup> century (e.g. Caneletto and Piranesi) use an eye height of about 2.5 times normal, which is about the extreme of its utility without loss of object identity. Bertamini, Yang, and Proffitt (1998) and Dixon, Wraga, Proffitt, and Williams (2000) have shown that one begins to lose the impression of object size when eye heights exceed this value. Interestingly, this is roughly the height typically attainable by raising a camera on a crane, a common device at the end of a film indicating that the film is over. Also, the opening shot and several others in Orson

Welles's 1958 film *Touch of Evil* use a crane effectively to dodge up and down within this range. Finally, partly because very high camera positions can defeat our sense of true object size, Hitchcock and others were able to use small models to film what would appear to be outdoor scenes.

Personal space immediately surrounds the observer's head, generally within arm's reach and slightly beyond. Within this region I claim five sources of information are generally effective (Cutting & Vishton, 1995)—occlusion and relative size from the linear perspective set, plus the reflexive, biologically engrained set of accommodation, binocular disparities, and convergence. Given that the latter two are not attained in standard film their absence could create a problem. Fortunately, the personal space of the viewer is not often relevant to film. Indeed, I claim that part of being a viewer of the action in a film is contingent on *not* having things enter one's personal space.<sup>4</sup> This impinges on one's person, and typically one does not want to be made aware of oneself when watching a movie. If you "lost it at the movies," to use Pauline Kael's (1965) felicitous phrase, you did so because you were not made aware of yourself. This is critical to Hollywood style.

Thus, whereas in the real world there appear to be three differentiable spaces (vista, action, and personal), in film there appears to be but one (action space). This makes the cinematographer's job possible. He or she doesn't have to worry too much about the background (indeed, many times sets can be substituted for outdoor scenes), and doesn't have to worry about the extreme foreground (because it would impinge on the space of the viewer).

### **How Cuts, Shots, and Narrative Knit Together a Film for a Perceiver**

Having broached spatial information and its use in cinema, let us turn next to temporal structure and how it interacts with space. It is useful to begin historically. Quite understandably, many early films were shot as theatre productions, with an unmoving camera in mid-audience. It was soon discovered, however, that the camera could move, execute close-ups, and the viewer could still make good sense of the action from different points of view. In addition, with increased demand and the advance of technology, films became longer and cuts were needed; one simply couldn't hold enough unexposed film in a magazine to shoot the whole movie (difficulties and expense of multiple takes aside). Early on, different shots were separated by a fading out of the first and then a fading in of the second. Darkness knits the two shots together. Later, dissolves entered the editor's toolkit, where the fading out of one scene is overlapped with the fading in of another (see Spotteswood, 1951). Nonetheless, quite early D.W. Griffith and others discovered that straight cuts were acceptable and not jarring (see Carey, 1982). Cuts separating shots in the same scene are by far the most common. In contemporary television within-scene, alternating shots and reverse shots account for more than 95% of all cuts (Messaris, 1994). Transitions separating shots from different scenes and times, however, often continued to use fades. For example, *The Sound of Music* in 1965 has straight cuts within scenes, but fades when both time and place are changed. More recently fades have passed out of favor, seeming quaint and unnecessary. A striking straight cut across scenes occurs early in Steven Spielberg's

1997 film *The Lost World*. An English family vacations on a remote, tropical island off Mexico. The daughter strays, and plays just off the beach beneath some palms. Small creatures surround her and attack. The scene then cuts to Ian Malcolm (Jeff Goldblum) with a palm tree behind him, but it turns out that Malcolm is in a subway in New York City and the palm tree is on an poster advertisement. Hollywood style is followed because the juxtaposition tells us that Malcolm will be connected to understanding the cause of the girl's death.

Why is a straight cut perceptually acceptable? This question divides several ways. First, why is it acceptable for one image to displace another taken from the same position in space, but with the camera rotated to a new orientation? Second, why is it *not* acceptable for one image to displace another taken from the same position and orientation? Third, why is it acceptable for one image to displace another taken from a different position with a new camera orientation?

**Cuts, saccades, suppression, and the lack of beta motion.** With respect to the first question, many conjectures have been made. In a 1965 interview, director John Huston made an intelligent start (see Messaris, 1994, p. 82), establishing himself as perhaps the first ecological film theorist:

All the things we have laboriously learned to do with film, were already part of the physiological and psychological experience of man before film was invented ... Move your eyes, quickly, from an object on one side of the room to an object on the other side. In a film you would use a cut. ... in moving your head from one side ... to the other, you briefly closed your eye.

Thus, for Huston, a cut is a surrogate for the real-world combination of saccade and blink.

Our visual world is usually continuous, but saccades and blinks do alter and cut the stream. We can often make ourselves aware of temporal discontinuities that occur during eye blinks, which are usually about a fifth of a second long (Pew & Rosenbaum, 1986). Make a blink longer than 200 ms and the “dimming” that often occurs becomes quite noticeable. Such dimming has a cause beyond mere lid closure. Some of the effect lies physiologically in the commands to the eyelid muscles. Such dimming occurs in the dark even when an optical fiber delivers light to the retina through the roof of the mouth (Volkman, Riggs, & Moore, 1980). Despite this, I know of no one who has collected normative data on the co-occurrence of saccades and blinks. This aside, it is quite clear that most of our saccades occur without blinks. So Huston's conjecture must reduce to one of comparing cuts with saccades.

Cuts are instantaneous, one frame to the next.<sup>5</sup> Saccade durations vary, mostly by the extent to which the eye moves, but 40 ms is about average, with a range of 20 to 90 ms (Hallett, 1986). The velocity of eye rotation during a saccade is quite fast with a range of 50 to 500°/s. Given that film screens are seldom seen wider than about 35° and that a full circle is 360°, this is fast indeed. During such movement one would expect to see blur. We do not. In fact, we see essentially nothing, a fact called saccadic suppression. Causes for this are complex but seem to be a mixture of blocking by two sources—feedback from eye muscle movements and a particular type of masking, or

blotting out of the message. Technically the latter is called metacontrast masking (Matin, 1986). In effect, we are relatively blind to visual information occurring from a few ms *before* a saccade, almost completely so during the 50 ms or so of a saccade, and tapering off for about another 50 to 100 ms *after* the saccade is complete (Volkman, Schick, & Riggs, 1968). The time course of interruption masking by metacontrast is about the same (without the presence of a saccade duration). Since interruption masking is likely to occur after a film cut, one can assume that we are partially blind to the visual information in the first 100 ms after a cut, about the duration of two frames. This means the editor must be a bit careful; one cannot cut quickly again. Quick cuts within this range are disruptive. They were tried, for example, in Dennis Hopper's 1969 film *Easy Rider*. Toward the end of the film, single-frame and longer shots were incrementally cut back and forth between scenes of motorcycle riding and camping. These were jarring, interfered with the narrative, and hence broke with Hollywood style.

However, masking and suppression explain only part of why cuts work. They explain a temporary blindness between shots at the cut line, and perhaps the lack of disorientation immediately after the cut, but they do not explain the acceptability of the cut. Why are we able to make sense out of two shots with no transition? Acceptability seems predicated, in part, on the physical differences between shots. Cuts become acceptable only when the general patterns of light in the two shots are sufficiently different (Hochberg & Brooks, 1996). This occurs naturally in fixations before and after a saccade. After rotating our eyes the backgrounds of what we see are different, the objects focused upon are typically different, the lighting is often different, and few edges and lines as projected on the retina line up across fixations. Thus, we accept a disrupted flow quite naturally, it is a part of our everyday visual world, and this is the heart of Huston's conjecture. But there is a caveat.

Unlike the perceiver in the real world, the editor composing the film must be careful with the content of successive images. If edges line up or worse, almost line up, a certain kind of irrelevant motion can occur, which I will call *beta* motion.<sup>6</sup> This motion occurs in the laboratory, and occasionally in neon street signs, where objects can change shape and position. This motion is not cinematic motion and would likely detract from the narrative. A particularly interesting candidate case occurs early in Stanley Kubrick's 1968 film *2001: A Space Odyssey*. Protohumans battle, one side wins, and a leader of the winning group tosses a bone into the air, which the camera follows and which rotates in slow motion. Cut to a spaceship docking at a space station. What is interesting about this cut is that across frames the bone and the spaceship do *not* line up. In fact, they are at right angles. Surely, the editor must have been tempted to align the orientations of bone and spaceship. Despite the fact that the backgrounds of the scenes are very different (light blue against the bone, and black against the spaceship), we can only assume that the editor found it inappropriate to align them; it must have created jarring *beta* motion.

The avoidance of *beta* motion is also part of the answer to the second question: Why is it not acceptable for one image to displace another taken from the same position and orientation? Juxtaposed shots taken from the same point of view create what is called a *jump cut*. Although used occasionally in French New Wave cinema of the mid-20<sup>th</sup> century, the perceptual effects of a jump cut are often very jarring. The

reasons for this would seem to be that the commonality of the backgrounds of the two shots across the cut anchors the sameness of what is seen (Hochberg & Brooks, 1996). Within this sameness, the changes in the focal object can often only be made sense of, perceptually, as plastic deformation and size change. Since people and cars and other objects that are the focus of the cinematic narrative cannot spontaneously deform or change size, anything indicating that they do seems weird, and detracts from the narrative.

**Shot-reverse-shot sequences, the cinematic viewer, and discontinuity.** The third question concerns cuts and the movement of the camera to a different position and orientation within the same scene. This is called a shot-reverse-shot pattern and occurs most often in filmed conversations. Such filming is a technical tour de force with great psychological interest. There are at least two interrelated problems here. First, after an establishing shot that shows two (or more) people in the scene, the camera typically frames each speaker sequentially, alternating position and focus between the two conversants. One person looks left off the screen as if to the other. The other individual looks offscreen right. Cinematic practice has shown that it is best if sight lines (gaze directions of the conversants) line up. The result is as if we (the camera) were a silent third party to the conversation, looking back and forth.

Because of the necessity of using long lenses, we (as the camera) cannot simply occupy a position nearby the conversants—they would have big noses. Thus, and revealing the second problem, we (as the camera) are often looking over-the-shoulder of one of the conversants, whether or not his or her shoulder is actually in the picture. In this manner, we do not occupy a single third position. How is it we can tolerate this subjective jumping around so much? My view is this jumping is only a problem if one assumes that we perceive the world metrically. Conversations are focused on the people, not the backgrounds, and we actually care quite little about the overall coherence of the background of the scene. We are perfectly happy, as viewers, if the camera positions are only roughly consistent with a third position generally between and to one side of the conversants. More simply, we are much more interested in the story than the details of the framework around those enacting the story.

This raises an important concern of filmmakers—continuity (see Anderson, 1996; Bordwell et al, 1985). Among other things, continuity means keeping track of what is in each shot and making sure that the world that is projected on the screen appears coherent. However, it appears that we, as perceivers, are not that particular about such coherence. Levin and Simons (1997; see also Simons & Levin, 1997), in one of the few laboratory experiments on the perception of film, showed that objects can appear and disappear across shots within the shot-reverse-shot sequence of a conversation, and viewers don't notice. In fact, one can sometimes switch actors across cuts as well (and actually do this in real life) without observers noticing. Indeed, Luis Buñuel did this with his leading actresses in his 1977 film *That obscure object of desire* and viewers often did not notice (see Buñuel, 1983).

However striking these results may seem, this kind of continuity “experiment” is forced by circumstances in some way upon the editors of nearly every film. In *The Sound of Music*, for example, consider a pivotal dramatic scene. Maria and the Von

Trapp children are having fun rowing on a lake behind the house. Captain Von Trapp suddenly arrives home with his betrothed, the Baroness Von Schrader, and the Captain excoriates Maria for a breach of strictness in the children's upbringing. Maria defends herself, to the point of being fired. The scene demands that it be shot outdoors, and that Maria and the children get wet, falling in the water. The clothes, being sewn from colorful drapes, are not easily replaced. Thus, different takes of the same scene could not be shot on the same day. The clothes must dry. The film version of the scene was clearly edited from shots taken on at least two days with quite different weather, one with clear blue sky and one with heavy humidity. The shots cut back and forth between clear and humid days seven times in the course of the argument between Maria and the Captain. Nonetheless, no student to whom I have shown this clip has ever noticed this fact, even after hearing a short lecture about continuity in film. Clearly, the narrative is sufficiently powerful that it doesn't matter that the sky behind Captain Von Trapp (taking up as much as half of the surface area of the screen) changes so many times.

### **The Filmmaker's Contract with the Viewer**

But there are constraints; not everything goes. This idea divides two ways. First, filmmaking demands that continuity be as great as possible during the actual filming process. This is because of the filmmaker's and editor's inability to know, in advance, which discontinuities would and would not be noticed—and psychological theorists don't know either. Unprepared for and noticeable discontinuity could jeopardize Hollywood style, and the success of the film. Second, whereas certain structural aspects of continuity may be violable, thematic aspects cannot. This raises the issue of montage, and the oft-described *Kuleshov effect* (e.g. Levaco, 1974; Pudovkin, 1958). The Russian filmmaker V.I. Pudovkin, a student of Kuleshov, made several short movies (each of three shots) using the actor Ivan Mosjukhin. In the first movie, the first shot showed a close-up of the relatively expressionless face of the actor, the second a coffin in which lay a dead woman, and the third another close-up of the actor. In a second film, the first and third shots were the same, but the second was replaced with a bowl of soup. Reports suggest that viewers read the expression on Mosjukhin's face in the third differently in the two short sequences. Such, it is said, is the power of montage. Indeed, Hitchcock embraced this idea (Truffaut, 1983, p. 216) and claims to have used it in his 1954 film *Rear Window* where, as a temporary invalid, L.B. "Jeff" Jeffries (James Stewart) views the murder of a neighbor across a back courtyard.

Yet there is much less here than actually meets the eye. What this description of montage leaves out is context. The montage will work, but only in the context of the longer narrative. Without that context, every experiment I know trying to replicate the Kuleshov effect has failed. Hochberg and Brooks (1996, p. 265) explain why:

Despite Eisenstein's assertion (1949) that two pieces of film of any kind, when placed in juxtaposition, inevitably combine into a new concept of quality, there is no reason to believe that without specific effort at construal by the viewer anything other than a meaningless flight of visual fragments ... will be perceived.

In other words, the filmmaker must first win over the viewer with the narrative. After the viewer accepts the narrative, the filmmaker has an implicit contract with the viewer to promote the narrative in an appropriate way. All storytellers in all media have such contracts with their audiences (see also Proffitt, 1976; Willats, 1995). At this point montage is good film practice, but only so long as the narrative continues in a satisfactory way. If it does not, the filmmaker has broken the contract, and the perceiver is on his or her own.

One final point about the acceptability of successive shots. Great importance in the psychological and film literature has been given to what is often called the 180° rule (see Carroll, 1980, Hochberg & Brooks, 1996, Kraft, 1987). This rule states that successive shots should not cross the line of sight between two conversants, or cross the line of action, but roughly any camera position within the remaining 180° is fine. Nonetheless, this rule seems violated quite often with little effect. In John Ford's 1939 film *Stagecoach* the opening scene cuts across the line of action (the stagecoach enters from the left, facing right, and we then see it facing in the opposite direction). Later, when more passengers are added, this "error" occurs again in the opposite direction, yet little seems lost. Few students, when shown the film, notice it.

More potently, the final scene in *Casablanca* cuts in violation of this rule during a three-way conversation among Rick, Victor, and Ilsa. Most of the conversation consists of shots cut between Rick (on the right) and Victor (on the left). Ilsa is between them, but closer to Rick. At a critical moment, however, suddenly Rick is on the left and Victor and Ilsa on the right. This is important for the story line, because at this moment Victor puts the papers of transit in his coat pocket, a gesture that could not be seen from the previous perspective. Moreover, the new placement of the three seals the fact that Ilsa is going with Victor, and not staying with Rick. Quickly, the camera positions shift back to Rick on the right and Victor on the left before departure. My experience in showing this sequence in a class is that no one is confused—indeed, no one even notices. This is probably because the positions of all three characters were well established in previous shots.

I agree with Murch (1995, p. 18), who suggests that the 180° rule is less important than often suggested, and is subordinate to many other purposes of film. I contend that such cinematic "rules" are not, as often proposed, like a "grammar" of film. In linguistics, violations of grammatical rules render sentences incomprehensible or ambiguous; in film, violation of these "rules" typically do not yield unknowable or uncertain results; one understands the film but is also aware that something is amiss. Instead of being like grammar, I contend these "rules" are like conversational axioms (Grice, 1957), the basis of a contract about how people behave towards one another in how they conduct a conversation. In film, these are parts of the contract between filmmaker and viewers. As I suggested earlier, the Hollywood-style contract dictates that filmmakers will not let viewers become aware of themselves. Crossing between conversants in the real world would be bad manners, and one would become aware of oneself. Crossing between conversants in film, as one apparently would in any violation of the 180° rule, would be equally rude.<sup>7</sup>

## Final Notes on Cuts and Time

Two final comments about cuts. First, can a film have no cuts? I know of only two in standard-release cinema, Hitchcock's 1948 film *Rope* and Louis Malle's 1981 film *My Dinner With André*. The latter actually has a beginning and ending shot outside a restaurant, but the 105 minutes in between is one 16-mm film shot (made with an extremely large film cartridge), occasionally with a gradual zoom in and out, of a dinner conversation. It is a remarkable film, but as a viewer one is teased and made aware of many things throughout. *Rope* is different; 80 minutes long, it is composed in 35-mm film as if it is one shot. As suggested earlier, in the context of describing a scene in *Notting Hill*, it is actually shot in 10 minute sections (see Truffaut, 1983). Breaks in the sections but not in the shot are hidden, for example, with slow pans across a person's back. Nonetheless, the action is continuous (walls and furniture having to be moved for the camera) making the film take place in real time. The camera roves throughout an apartment as a college professor (James Stewart) gradually discovers that two of his former students (Farley Grainger and John Dall) have followed the principles espoused in his course to an unexpected extreme, killing another former student. I showed this film to my daughters, and they never noticed that it was filmed in only one shot. Thus, short shots and cuts are not necessary to film; but it is equally unnecessary for films to be a continuous shot.<sup>8</sup> Our perceptual and cognitive systems accept either with equal alacrity as long as the narrative carries one's interest.

Finally, although films have had cuts and shots for a long time, it is clear that in recent cinema their pace is accelerating. Why? Many would blame music videos. The pace of shots and cuts in these 3-min clips can often be breathtaking, although the music overlay is continuous. Gleick (1999) would suppose that this pacing effect is a cultural one, due to the acceleration of demands on our time and a decrease in threshold for boredom. Indeed, to a degree, this is almost certainly true.<sup>8</sup> However, shot length beyond a second or two is not a biologically constrained. Bordwell et al (1985) found shot lengths in Hollywood cinema between 1930 and 1960 to be between about 6 and 12 seconds; today it is probably only a bit less. Only the lower limit of shot duration is limited by our perceptual and cognitive systems being able to make sense of shot composition and continuity, and this limit for any relatively sustained visual art form probably has a mean of about 1 to 2 seconds, or so. This is still below the pace of most music videos. Thus, I would claim that music videos exploit a heretofore unused perceptual and cognitive niche in cinema construction. I would claim further that our perceptual and cognitive apparatus has always been able to accept such pacing, it is just that only recently has this ability been tested. Film in general may evolve to have an average shot of slightly shorter duration than at present, but this would be a statistical artifact of mixing relatively long-duration shots (which in filmed conversations and elsewhere are not likely to diminish in length) with more music video-like shots.

## References

- Anderson, J.D. (1996). *The reality of illusion*. Carbondale, IL: Southern Illinois University Press.
- Atchley, P., Kramer, A.F., Andersen, G.J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychonomic Bulletin & Review*, 4, 524-529.
- Bertamini, M., Yang, T.L., & Proffitt, D.R. (1998). Relative size perception at a distance is best done at eye level. *Perception & Psychophysics*, 60, 673-682.
- Bordwell, D., Staiger, J. & Thompson, K. (1985). *The classical Hollywood cinema: Film style & mode of production to 1960*. New York: Columbia University Press.
- Buñuel, L. (1983). *My last sigh*. New York: Knopf.
- Carey, J. (1982). Convention and meaning in film. In S. Thomas (ed.) *Film culture: Explorations of cinema in its social context*. (pp. 110-125). Metuchen, NJ: Scarecrow Press.
- Carroll, J.M. (1980). *Toward a structural psychology of cinema*. The Hague: Mouton.
- Cutting, J.E. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, and Computers*, 29, 27-36.
- Cutting, J.E. & Millard, R.T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, 113, 198-216.
- Cutting, J.E. & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (eds.). *Perception of space and motion* (pp. 69-117). San Diego: Academic Press.
- Dixon, M.W., Wraga, M., Proffitt, D.R., & Williams, G.C. (2000). Eye height scaling of absolute size in immersive and nonimmersive displays. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 582-593.
- Eisenstein, S. (1970). *Notes of a film director*. Republished and translated from the 1948 Russian edition. New York: Dover.
- Eisenstein, S. (1949). *Film form*. New York: Harcourt.
- Gibson, J.J. (1950). *Perception of the visual world*. Boston: Houghton Mifflin.
- Gleick, J. (1999). *Faster*. New York: Pantheon.
- Grice, H.P. (1957). Meaning. *Philosophical Review*, 66, 377-388.
- Hallett, P.E. (1986). Eye movements. In K.R. Boff, L. Kaufman, & J.P. Thomas (eds.) *Handbook of perception and performance*, Vol. 1: pp. 10:1-112. New York: Wiley.
- Hochberg, J. (1978). *Perception* (2<sup>nd</sup> ed.). Englewood Cliffs, NJ: Prentice Hall.
- Hochberg, J. & Brooks, V. (1996). The perception of motion pictures. In M.P. Friedman & E.C. Carterette (eds.) *Cognitive ecology* (pp. 205-292). San Diego: Academic Press.
- Kael, P. (1965). *I lost it at the movies*. Boston: Little, Brown.
- Koenderink, J.J., van Doorn, A., & Kappers, A.M.L (1994). On so-called paradoxical monocular stereoscopy. *Perception*, 23, 583-594.
- Kraft, R.N. (1987). Rules and strategies of visual narratives. *Perceptual and Motor Skills*, 64, 3-14.
- Kubovy, M. (1986). *The psychology of perspective and Renaissance art*. Cambridge, UK: Cambridge University Press.
- Levaco, R., ed. & trans. (1974). *Kuleshov on film*. Berkeley, CA: University of California Press.
- Levin, D.T. & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Bulletin of the Psychonomic Society*, 4, 501-506.
- Lumet, S. (1995). *Making movies*. New York: Vintage.

- Matin, L. (1986). Visual localization and eye movements. In K.R. Boff, L. Kaufman, & J.P. Thomas (eds.) *Handbook of perception and performance*, Vol. 1, pp. 20: 1-45. New York: Wiley.
- Messaris, P. (1994). *Visual literacy: Image, mind, & reality*. Boulder, CO: Westview Press.
- Murch, W. (1995). *In the blink of an eye*. Los Angeles: Silman-James Press.
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually-specified events. *Cognitive Psychology*, 7, 480-494.
- Palmer, S.E. (1999). *Vision science: From photons to phenomenology*. Cambridge, MA: MIT Press.
- Pew, R.W. & Rosenbaum, D.A. (1986). Human movement control: Computation, representation, and implementation. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, & R.D. Luce (eds.) *Stevens' handbook of experimental psychology* (pp. 473-509). New York: Wiley.
- Proffitt, D.R. (1976). *Demonstrations to investigate the meaning of everyday experience*. Ph.D. dissertation, The State University of Pennsylvania. University Microfilms 76-29,667.
- Pudovkin, V.I. (1958). *Film technique and film acting*. I. Montagu, trans. London: Vision Press.
- Reisz, K. & Millar, G. (1968). *The technique of film editing* (2<sup>nd</sup> ed.) New York: Hastings House.
- Richter, J.P. (1883). *The notebooks of Leonardo da Vinci*. New York: Dover, reprinted 1970.
- Scharf, A (1968). *Art and photography*. London: Penguin Press.
- Simons, D.J & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1, 261-267.
- Sperling, G. (1976). Movement perception in computer driven visual displays. *Behavioral Research Methods and Instrumentation*, 8, 224-230.
- Spotteswood, R. (1951). *Film and its techniques*. Berkeley, CA: University of California Press.
- Swedlund, C. (1981). *Photography*. Fort Worth: Harcourt Brace Jovanovich College Publishers.
- Toulet, E. (1988). *Cinématographie: Invention du siècle*. Paris: Découvertes Gallimard.
- Truffaut, F. (1983). *Hitchcock* (rev. ed.) New York: Touchstone.
- Volkman, F.C., Riggs, L.A., & Moore, R.K. (1980). Eyeblinks and visual suppression. *Science*, 207, 900-901.
- Volkman, F.C., Schick, A.M.L., & Riggs, L.A. (1968). Time course of visual inhibition during voluntary saccades. *Journal of the Optical Society of America*, 58, 562-569.
- Wallach, H. & Karsh, E.B. (1963). The modification of stereoscopic depth-perception. *American Journal of Psychology*, 76, 429-435.
- Willats, J. (1995). The draughtsman's contract: How an artist creates an image. In H. Barlow, C. Blakemore, & M. Weston-Smith (eds.). *Images and understanding* (pp. 235-254). Cambridge, UK: Cambridge University Press.

## Acknowledgements

I thank Claudia Lazzaro for listening to near-endless ruminations about film during preparation of this chapter; Joseph Anderson, Barbara Andersen, and Dan Levin for comments on a draft of this chapter and for sharing their expertise about film, Dennis Proffitt for long-ago discussions of contracts; Michael Kubovy for discussions of Hitchcock; and my children, who for many years forced me to watch many different films with them *many* times, allowing me to become aware of and interested in Hollywood style.

## Footnotes

<sup>1</sup> Two important points need to be made. First, not all movies made in Hollywood are uniformly in Hollywood style, nor are non-American movies necessary not made in Hollywood style. The use of the term Hollywood style is intended to evoke the commonality of presentation and narrative found in popular and classic films (see Bordwell et al, 1985). Second, many genres contrast with Hollywood style, and for many reasons. Documentaries and television sportscasts have a narrative of sorts, but they differ from Hollywood style in that they typically have very long duration shots and no point-of-view editing. Newscasts also have a kind of narrative, but they differ from Hollywood style by having long shots and by having people look directly into the camera, intentionally engaging the viewer with eye contact, as in conversation. Advertisements and political spots differ in that they typically have no real narrative; instead, they have a strong message that their crafters want you to remember. Music videos differ in that have many very quick cuts, a continuous music line, and a sequence of shots that often alternates between the singer(s) and a small plot that uses the song as narration. Television sitcoms differ, having fewer changes of scene, using a generally proscenium set, and canned laughter. Finally, much of the film corpus of Eisenstein, for example, can be taken as a part of a genre of cinema that is trying strongly to educate the viewer, where the juxtaposition of content across cuts is often intended to elucidate similarities and dissimilarities, forcing the viewer to make judgments about what is seen.

<sup>2</sup> By dolly in I mean that the camera physically rolls closer to the object, and by zoom out I mean that the lens length of the camera gets shorter (normally minifying the objects in the image and creating a wider field of view). Surprisingly, Hitchcock and Truffaut (1983, p. 246) misdescribe the scene as a “track-out combined with a forward zoom.” Emmerich and Devlin's 1998 film *Godzilla* provides another example, although there are many. When Godzilla is about to erupt through the pavement of Manhattan, the camera is on Nick Tatopoulos (Matthew Broderick) and the buildings on the streets of New York convulse around him during the combined dolly in/zoom out.

<sup>3</sup> Motion is, of course, the *raison d'être* of film. One of the first “feature” films was the 15-sec film by Louis Lumière, *L'Arrivée d'un train à La Ciotat* (Ciotat being a small town outside Marseille). This 1895 film cost 1 franc per viewing and was an immediate smash hit, seen by breathless thousands (Toulet, 1988). Accounting for inflation, it is remarkable that viewing this film cost about \$685/hr in early 21<sup>st</sup> century dollars. We may bemoan the cost of going to the cinema today, but today's price is quite reasonable compared to previous times.

<sup>4</sup> Let me make two additional points. First, it is also not good Hollywood style to have actors look directly into the camera. If looked at, one becomes self-aware. For example, this is done early in *The Sound of Music*. Maria, returning late to the convent and being excoriated by the nuns, looks into the camera and shrugs her shoulders. It seems quite amateurish and disruptive. Second, the camera's extreme close-ups of actors' faces and other body parts do not necessarily impinge on the viewer's personal space. Instead, since long lenses are typically used in such scenes, the optics are akin to looking through binoculars, giving one a more immediate look at something that is still rather far away. These shots are *not* mistaken for being on top of the actor. For example, contrast them with some of the compelling, computer graphics shots in Disney and Pixar's 1999 film *Toy Story 2*. Many are subjective shots from the points of view of toys looking at human beings. The humans, of course, optically loom large—but not as with a telescope, but because they really are very close—entering the personal space of the toys. But rather than becoming aware of ourselves, this is part of the narrative, showing us what it is like to be a toy.

<sup>5</sup> Of course, standard film is usually interrupted 72 times per second, each of 24 frames three times by an episcotister. This brings the flicker rate above the normal human threshold, which for a bright light is about 60 times per second. Continuity at this scale is achieved by exceeding the temporal resolving capacity of the system. See Anderson (1996, Chapter 4) for a good analysis.

<sup>6</sup> Beta motion is a kind of apparent motion. There are many kinds of apparent motion and much confusion in the literature. It is sometimes said that film presents apparent motion due to its stroboscopic presentation of frames. But stroboscopic motion is, neurophysiologically, no different than real world and typically entails using many separate and sequential displays; apparent motion is quite different, and considerably less compelling (Sperling, 1976). Sometimes this distinction has been called short-range (for stroboscopic) and long-range (for apparent) motion, but this distinction is sometimes difficult to maintain. I use the technical and historical term *beta* motion in an attempt to avoid confusion. See Palmer (1999, pp. 471-479) for a good analysis and presentation of the types of apparent motion.

<sup>7</sup> Extraordinary blocking and camera gymnastics, as seen in a conversation in Ivan Reitman's 1988 film *Twins* between Danny DeVito and Arnold Schwarzenegger, may diminish the appearance of this rudeness.

<sup>8</sup> Hitchcock runs into a problem in *Rope*. Typically, in Hollywood style, when an actor looks offscreen, the next shot is a subject one, showing us what the actor sees. Resolved to have no cuts, Hitchcock could not do this. Thus, when the professor (James Stewart) finds an extra hat in the closet, he drops his head in thought, and turns the inside of the hat towards the camera which zooms in and shows us the initials of the dead man. This is a break with Hollywood style, because Hitchcock could not use point-of-view editing and the viewer is denied the subjective shot of what the professor sees. The information had to be conveyed by other means, one that borders on making us aware of ourselves. Hitchcock has often played with the subjective shot that is supposed to follow an offscreen glance. In his 1963 film *The Birds* the protagonist (Tippy Hedrun) sits on a

bench and looks offscreen several times, interleaved with shots of more and more birds arriving on a schoolyard jungle gym. We might have assumed that she was looking at the birds, but later she turns around in horror to see them. However, an establishing shot at the beginning of the sequence showed her facing away. See Carroll (1980 and Messaris (1994) for more discussion of this scene.

<sup>8</sup>Indeed, pace within a film is important too, and Lumet (1995) suggests that shorter shots are necessary to build to a climax. In *Twelve Angry Men* fully half of the cuts come in the final third of the film.