Three Filmmaking Practices That Guide Our Attention
to Popular Cinema

James E. Cutting*,**

Department of Psychology, Uris Hall, Cornell University, 109 Tower Road, Ithaca, NY 14853–7601, USA

Received 5 July 2021; accepted 24 November 2021

Abstract
Popular movies are constructed to control our attention and guide our eye movements across the screen. Estimates of fixation locations were made by manually moving a cursor and clicking over frames at the beginnings and ends of more than 30,000 shots in 24 English-language movies. Results provide evidence for three general filmmaking practices in screen composition. The first and overriding practice is that filmmakers generally put the most important content—usually the center of a character’s face—slightly above the center of the screen. The second concerns two-person conversations, which account for about half of popular movie content. Dialogue shots alternate views of the speakers involved, and filmmakers generally place the conversants slightly to opposite sides of the midline. The third concerns all other shots. For those, filmmakers generally follow important content in one shot by similar content in the next shot on the same side of the vertical midline. The horizontal aspect of the first practice seems to follow from the nature of our field of view and vertical aspect from the relationship of heads to bodies depicted. The second practice derives from social norms and an image composition norm called nose room, and the third from the consideration of continuity and the speed of re-engaging attention.

Keywords
Attention, center bias, conversations, eye trace, focal points, gaze, inward bias, popular movies

* E-mail: james.cutting@cornell.edu
** ORCID: 0000-0002-7338-846X

© Koninklijke Brill NV, Leiden, 2021
DOI: 10.1163/22134913-bja10032
1. Eye Fixations, Pictures, and Movies

We have known for quite some time that eye fixations to pictures show great variability across viewers and tasks (Buswell, 1935; Yarbus, 1967; see also Andrews and Coppola, 1999; Dorr et al., 2010). Thus, it was something of a surprise to discover that, across and within shots of a movie, most viewers tend to look at the same places at the same times (Goldstein et al., 2007; Hasson et al., 2008; Smith, 2013; Võ et al., 2012). It became clear that filmmakers of popular cinema are very good at controlling the uniformity of viewers’ gaze, a phenomenon called attentional synchrony (Smith, 2012; Smith and Henderson, 2008). I will use the term focal point for the central locus of this gaze conformity (see also Kirtley, 2018). Strikingly, this focal-point consistency appears to be driven by low-level features of movies; cognitive processes seem to play little role (Loschky et al., 2015).

Sergei Eisenstein (1949, pp. 15–16), the Soviet filmmaker and film theorist, separated concerns of the mis-en-scène (the three-dimensional placement of characters and objects in front of the camera) from the mis-en-cadre (their two-dimensional placement within the frame). The purpose of this article is to show that there are at least three filmmaking practices about mise-en-cadre that concern the focal point at the beginning of a shot.

1.1. What Is a Focal Point?

The point of focal interest, or focal point, is the place on the image where most viewers are fixating, or toward which they will soon saccade. A priori, however, it is not always clear where this should be. Consider two consecutive frames from Schindler’s List (Spielberg et al., 1993). In the top panels of Fig. 1, Oskar Schindler (Liam Neeson) is on horseback on a hill above the Krakow ghetto watching Nazis round up Jews for relocation to concentration camps. Earlier and repeating shots of the ghetto show a crowd with a little girl dressed in a red coat, noted with the arrow. Her coat is salient (in the movie but not in Fig. 1) because it is the only color in the main part of the film.

To de-semanticize these images a bit, the central panels of Fig. 1 show filtered outlines and textures in the scene, and more importantly the center points. Schindler’s face is mostly to the right side of the image, and the little girl is in the lower right. Where should a viewer look in each panel? In the left-middle panel the focal point is obvious. The little girl takes up only a small part of the screen. In the right-middle panel, however, I have drawn an ellipse of possibilities. Where on Schindler’s face do movie viewers look?

Fortunately, there are data that provide help in determining viewers’ potential gaze. One example, shown in the upper panel of Fig. 2, comes from Võ et al. (2012). Across a two-minute video linked to their article are short interviews of 20 people, mostly as single individuals (see Note 1). Viewers watched the video while being monitored with an eye tracker. They looked at areas...
around the character’s eyes, nose, and mouth with the sound on. With the sound off a different group looked basically in the same places but a bit less at the mouth. In both cases, however, the gaze clustering was centered near the tip of the nose with few fixations outside the face (Note 2). The pattern seen in the top panel of Fig. 2 is typical of the data seen from the movie.

Shots of characters are scaled by convention according to the size of the head and body. Shot scales are a continuous variable but it is typical to group them into seven categories (Bordwell & Thompson, 2004; Cutting et al., 2012). These are schematized in the small panels at the bottom of Fig. 2. The

Figure 1. The upper panels are adjacent frames from *Schindler’s List* (Spielberg et al., 1993), showing Jakob Schindler (Liam Neeson) looking down on the Nazi roundup of Jews in the Krakow ghetto in Poland, and a little girl (indicated with an arrow) in the crowd. In the edge-filtered middle panels, the small black squares are the middle of the images. The ellipses suggest regions of interest, locations of potential focal points, in both frames. And the bottom panels enlarge the ellipses and show two hypothetical focal-points within each, and note the horizontal distances between them. The inverse of the horizontal distance between focal points was used to weight the mean estimates for focal-point positions.
image in the top panel is a medium closeup, one that shows a character’s full face and shoulder area. We know from Smith (2013) that gaze clustering is tightest for this type of shot. The clustering around the nose is not because of its importance but because, at this scale in a standard home-viewing situation, a single fixation on or near the nose can bring in salient information simultaneously from both the eyes and mouth. This is critical because we look to the eyes (Messinger et al., 2012; Schurgin et al., 2014) and the mouth (Schurgin et al., 2014) for emotional expression, we read lips to augment the speech that we hear (Erber, 1969), and we look to the shoulders (Aviezer et al., 2012) and teeth, if they are visible (Calvo and Nummenmaa, 2008), for the intensity of expression.

Tighter-scaled shots, closeups and extreme closeups, spread the eyes and mouth and remove them from a single effective fixation area, as would be the
case of Schindler in Fig. 1. Longer-scaled shots – medium shots out to extreme long shots – bring other objects into view that might capture temporary interest of the viewer, but they also make the emotional expression harder to discern (Cutting and Armstrong, 2016). Both extremes of shots, shorter-scaled and longer-scaled, yield looser gaze clustering (Smith, 2013), but the face is still paramount in all shots with characters in them.

How should one determine where viewers look at beginnings and ends of shots? One method, common to many studies, is to use a group of subjects and track their eye movements while they watch a few short clips of movies. These data are excellent to have. The generalizability of such results to the average viewer is very good, but the generalizability of the clips to movies and to all of their shots is unknown. We do not know if the clips are representative. Thus, I chose another way to explore and to supplement our knowledge of the composition of focal points in popular movies.

2. Methods

2.1. Focal-Point Aggregation

There will always be some ambiguity, however minor, in determining a potential focal point of the beginning and end of a given shot – an ambiguity shared by the researcher with an eye tracker, the viewer, and the filmmaker who constructed the image. Moreover, despite early enthusiasm for saliency models (Itti et al., 1998), it seemed unlikely that any algorithmic image measurement would provide data of much use (Koehler et al., 2014). Thus, I twice went through 24 movies to estimate focal points in 31,254 shots, both at their beginnings and ends, marking in total almost 125,000 judgments on movie frames (Note 3). The general idea is to overwhelm with data the ambiguity of focal-point locations.

2.2. Movie Selection

I used 24 popular movies, each with a different director, that my students and I had previously analyzed. These are listed in Table 1. They are of three main genres – drama, comedy, or action movie. One of each was released in the eight calendar years from 1940 to 2010 evenly divisible by ten. They had a mean of 1,273 shots (range 295 to 2,756), and average shot duration of 7.3 s (range = 3.0–20.9 s). My students and I had previously measured many attributes of these movies (see Cutting, 2014, 2021, Cutting and Candan, 2015, and Cutting et al., 2012). They were released in one of five aspect ratios – the image width divided by its height. Six movies, those from 1940 and 1950, have ratios of 1.37 (presented at a resolution of 750 × 512 pixels), one from 1960 has a ratio of 1.66, eight from all years between 1970 and 2010 have a ratio of 1.85, one
from 1960 has a ratio of 2.2, and eight from all years between 1960 and 2010 have a ratio of 2.35 (1203 × 512 pixels, Note 4).

2.3. Conversation Shots and Other Shots

People chronically look at the faces of other people (Birmingham and Kingstone, 2009; Treuting, 2006), and 90% of all shots in popular movies have the face of at least one character in them (Cutting, 2015; Cutting and Candan,
2015). Of course, not all shots are medium closeups, as seen in the top panel of Fig. 2, but with medium shots (showing the character from waist up) these two shot scales are the most common in contemporary movies (Cutting, 2015, 2021). I will use the numerical codes listed in the bottom panels of Fig. 2 as a continuous scale in further analyses. These scales also generalize to shots without characters based on sizes analogous to the human body; a long shot for a car that nearly fills the screen, and an extreme closeup for a shot of the face of a cell phone. The mean shot scale for each of the 24 movies is shown in Table 1.

Cutting and Candan (2015) classified all the shots of these 24 movies into 15 exclusive categories and assessed their relative frequencies. These are shown in Table 2. In a plurality of shots, there is only one character on screen or only one whose face can be seen well (Note 5). Most of these are shots out of shot/reverse-shot clusters (an alternating mode of showing a dialogue, and with the character in view talking), or with two people in view but with only the face of the one seen talking. And sometimes the one person well seen in a

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Shot types, their definitions, and their relative frequencies.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shot types</strong></td>
<td><strong>Frequency</strong></td>
</tr>
<tr>
<td>Conversation shots</td>
<td></td>
</tr>
<tr>
<td>Shot/reverse-shots (with one speaking character)</td>
<td>22%</td>
</tr>
<tr>
<td>Over-the-shoulder shots (with one speaking character)</td>
<td>11%</td>
</tr>
<tr>
<td>Reaction shots (nonspeaking character being spoken to)</td>
<td>15%</td>
</tr>
<tr>
<td>Mediated shot/reverse shots (as on telephones, intercoms)</td>
<td>3%</td>
</tr>
<tr>
<td>Other shots</td>
<td></td>
</tr>
<tr>
<td>Singles (a stationary character not in conversation)</td>
<td>2%</td>
</tr>
<tr>
<td>Moving character shots (one or more characters, moving, speaking or not, often with a pan)</td>
<td>12%</td>
</tr>
<tr>
<td>Multiple character shots (two or more, stationary characters, speaking or not)</td>
<td>9%</td>
</tr>
<tr>
<td>Moving vehicle shots</td>
<td>2%</td>
</tr>
<tr>
<td>Action shots (extensive motion, typically covered with music, in a long sequence; 30% in action movies)</td>
<td>12%</td>
</tr>
<tr>
<td>Inserts (a shot of a detail within an ongoing scene)</td>
<td>4%</td>
</tr>
<tr>
<td>Cutaways (a shot taking the viewer away from an ongoing scene)</td>
<td>2%</td>
</tr>
<tr>
<td>Point-of-view shots (often without characters in them, following or coming before a shot with a character looking offscreen)</td>
<td>2%</td>
</tr>
<tr>
<td>Establishing shots (typically at the beginning of a new scene in a new location, typically long-scaled)</td>
<td>3%</td>
</tr>
<tr>
<td>Montage shots (a group of three or more related shots often linked by dissolves, typically covered with music)</td>
<td>2%</td>
</tr>
<tr>
<td>Combination shots (combinations of two or more of the shots above, and typically long in duration)</td>
<td>3%</td>
</tr>
</tbody>
</table>
dialogue is silent in what is called a reaction shot. For all of these, following from the data of Võ et al. (2012; see the top panel of Fig. 2), I chose the tip of the nose of the character as a focal point. One factor important to this analysis is that these shots are paired – one shot followed by another of this alternating type. I call these the Conversation shots.

Many of the other shots also have talking characters. These include those with multiple characters (two or more not in alternating shots, up to a crowd). In these cases, usually only one is person talking and I chose that individual’s nose as the focal point. Moreover, that character is most likely in focus and the other character(s) somewhat blurred, assisting in the choice. In certain singles (a shot with only one character in view) a stationary character is not in a conversation, and sometimes a single character is moving, talking or not. His or her nose was also used. Some shots are focused on moving vehicles and the middle of the windshield became the focal point. Point-of-view shots, inserts, and cutaways are generally focused on single objects, and their centers were used as the focal point. More difficult cases included action shots, for which still frames are often blurred. If a face was in the image it was used for the focal point, as above. Other difficult cases are shots of the wider environment and establishing shots, particularly those without characters in them. Here, I chose as a focal point what seemed to represent the locus of why the shot appeared in the movie. Together and simply, I will call all of these the Other shots. They are certainly a heterogeneous group, but they can provide a contrast with Conversation shots in that none alternate in pairs.

2.4. Task

I wrote a MATLAB script to go through each of these movies one shot at a time, taking two focal positions from each shot – one from the second frame and one from the second-to-last frame. The beginning frame of each shot had been determined by previous research (Cutting et al., 2010, 2012). Elaborating from Smith (2013), I will call the location of the estimated focal point in the early frame the entrance position and the location in the late frame the exit position.

For measurement purposes I moved a cross-shaped cursor to the spatial location of the tip of the nose of the most prominent character in the shot, or to the base of their nose if seen from the side, and clicked on that spot. For shots without characters I moved the cursor to what I felt was the focal point of the image frame, and clicked there. Each spatial position was digitally recorded and the next image automatically presented. Except for the image in the frame there were no other markings on the screen. Measured horizontally, my viewing angle was 20° wide and 8.5° high for the 2.35 movies, about the same as normal HDTV and laptop viewing.
I went through the 24 movies in pseudo-random order, performing the task on each whole movie before going on to the next (the tests). Then, a month after beginning, I started it all again (the retests) going through the movies in a different order. Each movie took from about forty minutes to six hours to go through.

2.5. A Consistency Check and Test–Retest Weighting

Figure 3 shows scatterplots of test and retest results for horizontal and vertical coordinates. The general concordance seems reasonable, with high correlations in both cases ($r = 0.85$ for the horizontal positions, and $r = 0.80$ for the vertical positions). Horizontal positions are normalized by aspect ratio (dividing pixel values by 1.37, 1.66, 1.85, 2.2 or 2.35) depending on the movie in which they appear.

I then averaged the test and retest coordinates at the beginning of every shot and also at the end of every shot. And I did this for all 31,254 shots. Next, I weighted those mean focal locations in two ways – by the inverse of the test–retest horizontal difference and the inverse of the test–retest vertical difference for the exit positions and the entrance positions. Examples of hypothetical horizontal comparisons are shown in the bottom panels of Fig. 1. This procedure dictates that when test and retest locations for a given focal point are close along one dimension, as in the lower-left panel, they receive more weight. When test–retest locations are farther apart, as in the lower-right panel, they receive less weight.

Figure 3. Scatterplots of test–retest concordances of judged focal points in both horizontal and vertical image dimensions. The data for the horizontal dimension are normalized by the aspect ratio of the movie in which they appear.
3. Results

3.1. Raw Data

Aggregated focal-point positions are shown in Fig. 4. These scatterplots represent the 22 movies with the three most common aspect ratios and the beginning focal points (entrance positions) of all shots. The data for the end positions (exits), and their statistics, are virtually identical. There were no effects among genres or release years.

Two features are apparent in Fig. 4. First, the mean within-movie horizontal spread of focal points away from the vertical midline is strongly correlated...
with the aspect ratio of the movies \( t(19) = 4.07, p = 0.0007, \text{Cohen’s } d = 1.87 \). However, once the data are normalized, dividing each mean horizontal spread by the aspect ratio, there is no pattern of aspect-ratio differences \( t(19) = -0.37, p = 0.71 \). Thus, filmmakers control gaze within an ellipse affinely stretched by the aspect ratio. Given this normalized null result, I report many distances below (whether horizontal or vertical) in terms of relative height of the image, which was constant in the presentation of all movies. This also renders results independent of image size.

### 3.2. Elevated Centering of Focal Points

The second apparent feature of Fig. 4 is that the centroids of the data are elevated above the middle of the image, as in the gaze cluster in the top panel of Fig. 2. In particular, assessing the mean vertical position in each movie revealed that, compared against the center of the screen, the centroid of focal points is located 10.5\% [standard deviation (sd) = 3.2\%] of screen height above that center [one-sample \( t(23) = 16.3, p < 0.0001, d = 3.3 \)]. Indeed, only 15\% of all focal points fall below the center. This effect was also seen by Cutting (2015) for a much smaller sample of character-only locations and by Breeden and Hanrahan (2017) for eye fixations across a varied collection of shots in other popular movies. It seemed due to the fact that characters’ heads are on bodies and, when more than just the head is in view, the body pushes the head upward in the image. If so, this result is due to differences in shot scale.

To assess a possible effect of shot scales I used the data from Cutting et al. (2012) for all shots in these same movies. Here, the vertical location of the focal point against shot scale and in Conversation shots predicted the elevation of focal points \( t(15,257) = 14.97, p < 0.0001, d = 1.19 \), the means of all of which are above screen center. The overall pattern is shown in the scatterplots in the left panel of Fig. 5. In particular, when a character is shown in the image in an extreme closeup, the nose is generally closer to the center, and nose position generally grades upwards in the image with longer shot scales (although there is no difference among medium long shots, long shots, and extreme long shots; Note 6).

For the Other shots there was no shot-scale pattern \( t(15,591) = 0.62, p = 0.53 \), as suggested in the right panel of Fig. 5. Mean focal position of all such shots was equally elevated above the center of the screen. Importantly, in a linear mixed-model regression with movies as a random variable, I assessed the elevation differences of focal points across shot scales in Conversation shots against the general elevation uniformity of those in the Other shots, and found a Shot Scale × Shot Class (Conversation/Other) interaction \( t(30,038) = 129, p < 0.0001, d = 1.49 \).
In addition, there is a subset of the Other shots whose results are also pertinent. Most Other shots have people in them at different shot scales, so it is not surprising that focal points for these shots are generally elevated. But some classes of Other shots rarely have characters’ faces in them. They include point-of-view shots (a shot of what a character is looking at after or before looking offscreen), inserts (a shot, blown up, typically of a detail within a previous shot), and cutaways (a shot that departs from the main action in a scene, say an alleyway dialogue, to something thematically relevant, like a scampering black cat; Note 7). These are italicized in Table 2. There would seem to be no non-contextual reason to elevate these shots above the center of the screen, but they nonetheless appear 8.1% above it [sd = 11.1%, \( t(2,263) = 34.5, p < 0.0001, d = 0.72 \)].

### 3.3. Exit–Entrance versus Entrance–Exit Focal-Point Distances

Again, the entrance position is the focal point at the beginning of a shot; the exit position is that at the end of a shot. Thus, exit–entrance distances are those
for which eye movement excursions would be made across shots; entrance–exit distances would be those made during the shot (or more strictly, since no scanpath need be linear, the linear difference between opening and closing focal-point positions).

Scaled to the height of the images (again, presented at 512 pixels and 8.5° in height) the raw median, Euclidean, exit–entrance (across-shot) distances were 26.7% of screen height (mean = 31.3%, sd = 32.2%). The raw median, Euclidean, entrance–exit (within-shot) distances were 10.7% of screen height (mean = 18.1%, sd = 20.0%). Thus, the within-shot focal point differences were only about 40% as large as the across-shot differences [one-sample \( t(31,238) = 83.3, p < 0.0001, d = 0.47; \) Note 8]. In a similar vein, the correlation for horizontal and for vertical coordinates at the beginnings and ends of the same shots are relatively high (\( rs = 0.67 \) and 0.56, respectively) and higher than those for the ends of given shots and the beginnings of those that follow (\( rs = 0.30 \) and 0.26).

### 3.4. General Exit–Entrance Side-Switching for Conversations

Popular movie dialogues often begin with an establishing shot, showing two characters on the opposite sides of the screen, facing one another, possibly with one having just walked into the scene. After establishing their positions, subsequent shots typically have one character slightly to one side of the screen facing to the center alternating with a shot of the other character slightly to the other side facing back to the first.

Across the Conversation-shot pairings, the proportion of focal points that switched sides was calculated separately for each of the 24 movies. In this manner, movies with more shots could not dominate those with fewer. Results showed that 55% of the focal points of those pairs switched sides \( (t(23) = \) \( -3.78, p < 0.001, d = 1.6) \). In a separate analysis, genre and release year were also considered. There were no differences across dramas, comedies, and action movies, but the proportion of side-switching pairs has increased over time \( [t(20) = 3.52, p = 0.006, d = 1.57] \), from only about 50% for movies from 1940 and 1950 to about 60% for those from 2000 and 2010, with the others generally in between. Results are shown in Fig. 6 (Note 9).

One might have thought that these results would be driven by aspect ratio, but they are not \( [t(20) = 1.06, p = 0.31] \). The reason appears to be that the use of cinemascope in the early 1950s trained filmmakers to exploit the lateral space of the image more and more regardless of the aspect ratio they were dealing with (Cutting, in press). The pattern shown in Fig. 6 is one of the many gradual changes in popular cinema over the last 70 to 100 years (Cutting, 2021).
Finally, after normalizing lateral dispersion by aspect ratio and treating movies as a random variable within this dispersion, a linear mixed-model regression revealed no effect across aspect ratios $F(1,22.8) = 0.05, p = 0.99$, with a mean lateral displacement from the midline of 7% of screen width, and 14% between alternating pairs. What this means is that conversants have been shown incrementally more separated with greater aspect ratios. This separation can also be used for narrative purposes, discussed below.

3.5. Less Exit–Entrance Side-Switching for Other Shots

The Other shots – which are also the other half of popular cinema – are a haphazard group with very little in common except that they are not typically parts of repeating pairs. Because of their diversity, one might expect somewhat more diffuse results in whatever one might measure. As with the Conversation shots, the measurement here concerns successive exit/entrance focal points that either switch sides or do not.

The Other shots were accumulated within each movie and again assessed in their percentages of switching sides. In contrast to the Conversation shots, successive focal points for the Other shots switched sides only 46% of the time $t(23) = 5.44, p < 0.0001, d = 2.27$. There were no effects of genre and release year. Again, after normalizing lateral dispersion by aspect ratio, and treating movies as a random variable within it, a linear mixed-model regression revealed no effect across aspect ratios $F(1,20.5) = 1.83, p = 0.19$. As a summary of the two classes of shots, Conversation versus Other, the
side-switching contrast between them is shown in Fig. 7 \( t(23) = 5.74, p < 0.0001, d = 2.39 \).

4. Discussion: Three Rationales in Constructing the *Mise-en-Cadre*

4.1. The Power of the Center, Sort of

The center of any rectangular frame has a strong attentional pull (Arnheim, 1988; Guidi and Palmer, 2015; Hubbard, 2018). Moreover, looking there would maximally fill the V1 cortical response areas for a rectangle of that size, and the larger the rectangle the greater the extent of the neural area stimulated. And in constructing their images, photographers (Tatler, 2007) and cinematographers (Tseng *et al.*, 2009) are said to place the most important information in pictures and shots near the center. Obligingly, viewers have a strong disposition to look there, called the center bias (Tatler, 2007). Moreover, in an analysis of viewers’ eye movements while watching movies the strongest tendency after a cut is for viewers to saccade back toward the center of the screen (Mital *et al.*, 2012).

But the results here and those for the fixation-position data of Breeden and Hanrahan (2017) and Võ *et al.* (2012) suggest that, while focal points tend to be centered horizontally, they are not centered vertically. Instead, they are about 10% above the center – call it an above-center bias. The reason is that, for most shots of characters, their heads are rarely centered vertically in an image. Placement depends on shot scale, or how much of a character’s body is shown. Faces carry most of the emotional information, but shoulders and bodies are not mute, and filmmakers often find it important to show them. The more of the body shown, the more the face will tend to be pushed upward in the image.
4.2. Turn-Taking, Retracted Space, and Nose Room

Conversations are about taking turns (Sacks et al., 1974; Schegloff, 2007). Beyond movies, dialogues are most common in real life as well. One person talks, then the other. Moreover, the alternating turns appear deeply engrained in our social norms. Infants are well aware of the turn-taking rules long before they can talk (Bloom et al., 1987). Movie viewers obviously cannot participate in movie dialogues, but filmmakers seem to have assumed, increasingly over release years, that the rule should be followed by the characters, likely because that is what happens in real life.

When watching a dialogue in the real world, we are accustomed to being able to see both people, one person to the right of us and the other to the left. This sets up a spatial opposition. Similarly, the two conversants typically face one another, which sets up the directional opposition, the one on the right facing left and the one on the left facing right. But we want to see their faces clearly. Because the conversational pair usually maintains social distance, we cannot see both clearly at the same time. This sets up separate, alternating single views (or, less often in popular cinema, alternating over-the-shoulder views). The space between the characters then, in topological terms, is retracted and, for the sake of the separate images, it no longer exists.

Moreover, the face of each character often needs a bit of frontal space — sometimes called nose room or look room for a stationary person, lead room for a moving person or object (May, 2006), and these are related to an inward bias more generally (Palmer et al., 2008). This displacement creates a type of negative space in front of the character — a space with psychological impact, but without important content (e.g., see Sante, 2008; Seitz, 1963) (Note 10). Thus, nose room pushes the characters laterally into spatial opposition.

This negative space can also be used to convey relationships. Consider the examples shown in Panels 1 and 2 in Fig. 8. I have chosen them because David Fincher, the director of The Social Network (2010), is known to be very concerned with the details of the mise-en-cadre (Note 11). In Panels 1 Erica Albright (Rooney Mara) and Mark Zuckerberg (Jesse Eisenberg) have a difficult discussion. They are shown in successive pairs of over-the-shoulder shots, and they are on far opposite sides of the screen with a lot of negative space between them. Why? Because they are arguing, and they are far apart in their opinions. Distance in the mise-en-cadre thus reflects distance between opinions. Indeed, they break up at the end of the scene, and in a tiff Zuckerberg goes back to his dorm and begins to blog and code the forerunner of Facebook. In Panels 2, outside a winter party, Eduardo Saverin (Andrew Garfield) and Zuckerberg are also on opposite sides of the screen in shot/reverse-shot singles, but not so far apart. Why? Because Zuckerberg is successfully convincing Saverin to advance him more money. Again, screen distance reflects social
distance. Although later in the story they will drift apart – and be shown farther apart while still on opposite sides of the screen – here they are still partners.

4.3. Eye Trace

An early description of filmmakers’ control of the mise-en-cadre comes from two English visitors to Hollywood, Gordon and Gordon (1930, pp. 105–106). They suggested considering three shots – A, B, and C. Shot B should:

“begin with the interest concentrated almost at the point which the eye was watching at the close of A, or it must continue a movement suggested by A. Again, whatever action may be that runs through B, … C must pick up the interest at the local spot where B ceases. … This has no reference to the story itself,
but merely to the making of the pictures considered only as spots of colour and centres of pictorial interest.”

This superimposition of focal points at the beginning of one shot over the end of the previous minimizes eye trace (Murch, 1995), the saccade length between shot-exit and shot-entrance positions (Smith, 2013).

Eye-trace minimization goes by several other names. Elsewhere in the film-editing literature this has been called the screen-position edit (Bowen and Thompson, 2013), or a continuum of movement across shots (Block, 2008); and in the human–computer interaction literature it has been called the collocation of objects of interest across a cut (May and Barnard, 1995; May et al., 2003). The assumption here is that the cut is an exogenous cue to prepare for an attention shift away from the end focal point of one shot, but if that focal point is replaced by new focal point in the same location the reengagement would require no location shift and attention shifting would be quicker (Talsma et al., 2011).

From the data here, there is reasonable evidence for eye-trace minimization from two sources. First, the Other-shot data show that focal points tend to stay on the same side of the screen. Second, focal points of point-of-view shots, inserts, and cutaways are above the center of the screen to almost the same extent as the Conversation shots. There would be no reason to place them there unless previous and subsequent shots in dialogues had elevated focal points as well. Given that these shots follow, precede, and are occasionally intermixed within conversations, their elevation makes sense.

Again, let us return to Fincher and The Social Network. Panels 3 and 4 of Fig. 8 show pairs of shots after Zuckerberg has returned to his dorm room and begins to blog and code. In many shots we watch his computer screen; in others we watch his fingers type. Panel 3a shows the end of a shot in which we watch the scroll of letters move across the screen left-to-right. Panel 3b shows the next frame. We find Dustin Moskovitz (Joseph Mazzello) light up a joint, very near to where we would be fixating had the scrolling type of the previous shot continued. Similarly in Panel 4a we find Zuckerberg blogging and we would be looking at his fingers. In the next frame, a point-of-view shot shown in Panel 4b, we find a new scroll of letters of what he is typing. In both cases, eye trace – and remaining in the same portion of the same side of the screen – is minimized.

5. Four Caveats and Replies

Despite the evidence in support of these three practices, some readers may have concerns. First, the corpus investigated contains only English-language movies. Might the movies of other national cinemas be different? Perhaps.
But, over the past decades, more than half the movies viewed worldwide are Hollywood movies (Note 12), so it seemed prudent to start with these.

The second concerns motion. The spirit of eye trace in the descriptions by Block (2008), Gordon and Gordon (1930), Murch (1995), and Smith (2012) is of a moving eye. In contrast, the data presented here are entirely about static points in screen space. Yes, it is likely that the focal point at the end of a shot and that at the beginning of the next are often both on the move, as in Fig. 8 from Panels 3a to 3b and from 4a to 4b. This would appear to be a problem, but in a spatially coarse, data-heavy situation like this one with motions likely in all directions, I think it is not a serious one. One the one hand, characters in dialogue do not move very much (as shown in the entrance–exit data) and, on the other, to the extent that some across-cut motion would cross the midline, the results here would underestimate the frequency of eye trace in Other shots.

A third issue concerns the marking of the focal points in shots using the tip of a character’s nose. Võ et al. (2012) eye-tracked viewers with and without hearing the soundtrack; I viewed these movies without a sound track. They noted that when a character is heard to speak the array of viewer focal points on the mouth increased from about 20 to 30%. For the person shown in Fig. 2 in medium closeup, the distance between her mouth and eyes is about 30% of the screen height. Assuming that the voice redistributes 10% of gazes from the mean of nose/eye area to the mouth, this would lower the centroid of fixations in the image about 4% of the screen height. But medium closeups in these movies account for only 15% of all shots in this sample. The mean shot scale in this movie sample is roughly a medium shot, showing the character from the waist up. This decreases the size of the face in the image by half or more, and would reduce the effect of more focus on the mouth to about 2% of screen height or less. Thus, this consideration is not large enough to nullify the elevation effect. Moreover, since this difference is measured vertically, it is irrelevant to the two screen-side effects.

And a final issue is that I as experimenter, Ebbinghaus-like, was also the subject. Biases are certainly possible, although the task here was quite simple and easy to execute automatically. In defense of my empirical strategy, let me emphasize that the results here align with four well-established empirical results and two minor guidelines for making movies. Among the results: first, faces are the dominant social objects looked at by people (e.g., Birmingham and Kingstone, 2009; Note 13), even by infants (Fantz, 1961; Frank et al., 2009). Second, popular cinema is rife with faces (e.g., Bolton, 2019), most often presented one at a time (Cutting and Candan, 2015). Third, faces are centered slightly above midscreen and viewers have a strong tendency to look there (Breeden and Hanrahan, 2017; Võ et al., 2012). And fourth, viewers tend to look near the nose of a character (Võ et al., 2012), extracting emotional information from the eyes and mouth (Schurgin et al., 2014; Smith, 2013).
One contribution of this article is to link all four of these established results, while qualifying the decontextualized view that the center of a rectangle draws attention (Arnheim, 1988; Guidi and Palmer, 2015; Hubbard, 2018) (Note 14). Thus, given this confluence, that I was both experimenter and subject seems less problematic.

Concerning relevant filmmaking guidelines, it is occasionally reported about filming dialogues that they should be staged by putting the two characters on opposite sides of the screen (Note 15), although it has surprised me that I could find so little discussion of this. And among the ‘rules’ for composing the relationship between the end of one shot and the beginning of the next, it has been said that the filmmaker should align the focal points as best one can – although the major contemporary proponent of this idea, Academy Award-winning editor Walter Murch, suggests this rule accounts for only about 7% of the rationale for editing (Murch, 1995).

Yet these two guidelines are irreconcilable. Filmmakers cannot follow the focal point of a shot that is on one side of the screen with that of the next on the same side if they are also placing conversants on opposite sides. Moreover, insofar as I have been able to discern, this incompatibility has never been addressed. A second contribution of this article is to separate the regime of dialogues, where side-opposition is prominent, from all other shots where evidence for eye trace can be found. In this manner the results here provide empirical support and psychological rationale for both guidelines, but in different and constrained contexts.

And a final contribution is that previous eye-movement research with movies has generally used short clips taken from larger works. As important as these results are, we do not know how representative the clips used are of movies in general. In making 125,000 judgments from 24 whole movies released over seven decades – a scope that would be impractical in an eye-tracking environment – I suggest that the results reported here can be taken as representative of all of popular cinema, and to compliment previous eye-movement research.

6. Conclusion

On average, the people in Western countries spend as much as three hours a day watching edited video, and may watch as many as eighty movies a year across all platforms (Note 16). Over the last 20 years researchers have discovered that our collective eye-movement behavior when watching movies is quite different than that while perusing static pictures. Moreover, filmmakers go to considerable lengths to control our eye movements and they succeed to an astonishing extent. Thus, it seemed prudent to understand more about how they do it and what they actually do.
Together, the data presented here support the idea of at least three filmmaking practices for the composition of images on the screen, what Eisenstein (1949) called the mis-en-cadre. The first is that filmmakers place important content slightly above the screen center. This is where the center of a character’s face is mostly likely to be found. The reason for the horizontal centering would seem to be registering the screen within our field of view to maximize overall cortical responses. The reason for the slight elevation concerns the tradeoff of balancing the major sources of emotional information (a character’s eyes and mouth) with minor sources (the rest of the body).

The second practice concerns conversations, which are typically presented in a series of opposing pairs. Filmmakers alternate focal points of adjacent shots slightly to either side of the midline, and they seem to have done this more and more since the middle of the 20th century. This guideline follows from social norms and spatial considerations. The social norm is turn-taking. The spatial consideration is that the character’s face is not at midscreen; instead, there is nose room – a larger amount of screen space in front of the face than behind – pushing the face to one side of the image. This creates two general oppositions – facing direction and screen side (Cutting, in press).

The third practice applies to all shots external to dialogues, those that are not constructed in pairs. Here, filmmakers tend to place focal points of the ends and beginnings of successive shots to the same side of the vertical midline, roughly following the notion of eye trace. Use of this guideline minimizes saccade lengths, and can speed up viewer processing.

But who decides on these practices? No one. At least I think no one overtly and consciously decides on them as a movie is made. These practices and many others in the production of popular cinema have evolved over decades of experience (Cutting, 2021). One instance of that evolution can be seen in Fig. 6, where conversing characters have been gradually placed more often on opposite sides of the screen. Clearly, because there was no instantaneous change, the practice seems likely to have spread through a process like cultural exchange. As sound and film editor Walter Murch noted, “You pick up the good things that other editors are doing and you metabolize those approaches into what you are doing, and vice versa” (Ondaatje, 2002, p. 62). Directors, cinematographers, and script writers would likely agree concerning other directors, cinematographers, and script writers. It is all in the collective development of the craft.

Acknowledgements

This article was presented as a paper at the annual meeting (virtual) of the Society for Cognitive Studies of the Moving Image, June 2020. I thank Chris McManus, Tim J. Smith, and other reviewers for insightful comments. Data are available at http://cutting.psych.cornell.edu/data.htm.
Notes

1. For the research video see: https://vimeo.com/visualcognition. The original film, “Fifty people, one question: Brooklyn” is part of a series by Crush and Lovely. It is nearly 6 min long and can be found at https://www.youtube.com/watch?v=VJAUGg4081Q. Both accessed 3 July 2021.

2. Haensel et al. (2020) found some cultural differences in gaze in watching movies. A group of British viewers looked more at the mouth of a talker, and a Japanese group looked more at the eyes. This contrasts with the photographic results of Miellet et al. (2013), who found for recognition purposes that Westerners looked more at the eyes and mouth and Chinese observers looked more at the nose. In both cases, however, the nose would be the mean locus of all fixations.

3. This was a COVID-19 self-quarantined project. Also, at the time of this study MATLAB did not have a face-recognizing algorithm available that was adequate to the task.

4. Aspect ratios can be confusing. The Academy ratio is sometimes listed as 1.33. However, this is the modal ratio of movies before embedded sound, and before the standardization in 1932 of 1.37. All of the relevant movies used here were listed as the latter. In addition, cinemascope comes in several aspect ratios. The earlier movies in this sample have a ratio of 2.35 and those in 1990 and later have a ratio of 2.39 or 2.40. For simplicity’s sake, except for data analyses, I use 2.35 throughout this article.

5. Table 1 shows the average sampled number of characters on screen (Cutting, 1915) with the proviso that all group shots have their count truncated at 5.

6. It might seem likely that the overall elevation of focal points above the center of the screen is due to the fact that, in crowded movie theaters, viewers may have the lower part of the screen occluded by other viewers sitting in front of them, and that filmmakers know this. However, this account does not predict the data of elevation variation by shot scale.

7. What might be called a cutaway to a character was classified as a reaction shot; inserts were constrained to autonomous shots outside of conversations; and point-of-view shots to a stationary character are not common (Cutting and Candan, 2015).

8. Here and elsewhere the degrees of freedom differ from that garnered from total number of shots due to missing data. Here the missing data include the fact that the first shot has no previous-shot exit position, and the last shot has no subsequent entrance position.

9. The overall side-switching effect would be stronger had I used the center of the character’s head rather than the nose as the focal point. There is a
strong tendency for filmmakers to pose conversants not only on opposite sides, but also facing toward the middle of the screen (Cutting, in press). This staging moves the character’s nose toward, and sometimes past the center. However, this study concerns focal points and most viewers look near the nose, not the center of the head.

10. Negative space is a term important to discussions of Japanese gardens, to dance, and to art, particularly the work of Hans Hofmann. It is also relevant to figure–ground reversals.

11. For examples of Fincher’s technique, see: https://www.youtube.com/watch?v=F3ZSX3D1dUI; https://www.youtube.com/watch?v=QChWI Fi8fOY; and https://www.youtube.com/watch?v=agzKBtqVmLE.


14. See also McManus, Stöver, and Kim (2011). They pushed Arnheim’s ideas empirically and found only modest support for them.


References


Eisenstein, S. (1949). *Film Form: Essays in Film Theory*. (J. Layda, Transl. and Ed.), Harcourt, San Diego, CA, USA.


