

You Had My Curiosity, but Now You Have My Attention:¹

Camera Movement and the Spatial Distribution of Audience Attention in Cinema

As film and video content becomes increasingly ubiquitous, the importance of understanding the cognitive effects of film grows in equal measure. Simultaneously, filmmaking technology has improved at such a rate that filmmakers of all levels have the ability to perform virtually any camera movement or technique. However, the decisions about what filmmaking technique to use to capture a given image is typically based upon a series of unquestioned axioms that date back to the Golden Age of Cinema. This project will seek to scientifically explore one filmmaking convention and its impact on the allocation of audience attention. This study will examine how a camera dolly in or dolly out (also known as a push in and a pull back) affects the spatial distribution of visual attention. The answer to this question will be of special interest to filmmakers, as they can employ this information to make informed decisions about what visual and narrative information should be conveyed with a dolly shot and where within the frame that information should be delivered. Additionally, this research will open a line of quantitative inquiry into the qualitative effects of creative and aesthetic decisions made by filmmakers, leading to a better understanding of the way video content is made and watched.

Daniel Matyas

Yi-Chia Chen and Brian Scholl

Yale University

Department of Cognitive Science

¹ *Django unchained* [Motion picture]. (2013). Beverly Hills, CA: Anchor Bay Entertainment.

There are dozens of unique and interrelated techniques for creating a shot in a movie, ranging from the extremely simple, such as a low or high angle shot (FIGURE 1.1 and 1.2), to the highly complex and technical, like a 4 minute continuous take on a Steadicam (FIGURE 1.3-1.5). There are a variety of systems for categorizing shots that involve a moving camera, ranging from formalist (e.g. "There are seven basic moving camera shots..." [Giannetti 2011]) to functional (e.g. Nielsen 2007). Regardless of an individual's preferred system, nearly all film theorists claim that each technique or movement can be used to convey narrative information, and each has distinct perceptual effects on audiences (Eisenstein 1975). Here are a few filmmaking "rules" regarding composition and other creative decisions from a textbook on film theory:

A character moving towards the camera is seen as aggressive, movement away from the camera is reclusive.

Even numbered units (of characters, objects, etc.) are stable, odd numbered units are unstable.

High-angle camera placement creates a sense of confinement.

"Negative space" creates a feeling that something is missing for the audience.

Widescreen is used to create a feeling of expanse and size.

Giannetti, 2011

While anecdotal evidence and many experts in the field of Film Studies support these sorts of claims about filmmaking technique and its influence on audiences, there has been only sporadic scientific study of these effects. The most prominent contemporary psychological research program concerning narrative cinema focuses primarily on the visual and temporal aspects of existing films: how films are edited,

patterns in color schemes, the order of shots used to present a scene, and how events are segmented in movies, among others. However, the program's structure is almost entirely statistical. Using a library of 160 films, typically whittled down to 24 for close analysis, the researchers find statistical patterns and trends in blockbuster movies and compare them to existing conceptual models and theories of attention, as well as claims made by film critics about trends in cinema (Cutting et al. 2011; DeLong, Brunick & Cutting 2014; Cutting & Iricinschi 2015; Bordwell 1997; Thompson 1999).² While this research is interesting from the standpoint of film theory and statistical analysis, it is made up of purely observational studies that cannot assert causal relationships between the nature of the films and the nature of the human mind.



FIGURE 1.1 – High-Angle Shot, *The Avengers* (2012)



FIGURE 1.2 – Low-Angle Shot, *The Avengers* (2012)



FIGURE 1.3 – An early frame from a long continuous Steadicam shot, *Boogie Nights* (1997)



FIGURE 1.4 – A later frame from the same continuous Steadicam shot, *Boogie Nights* (1997)



FIGURE 1.5 – Behind the scene of a Steadicam shot, *Birdman* (2014)

² See http://people.psych.cornell.edu/~jec7/pictures_film.htm for a complete list of lab's publications.

Early experimental work on film and psychology aimed to prove that camera angle could convey narrative information. One study found that viewing characters from a low-angle caused participants to evaluate them as stronger, bolder and more aggressive, while a high-angle shot had the opposite effect on participants' perception of a character. These results were found through subjective ratings and forced choice responses (Kraft 1987). While these findings may be entirely valid, mere exposure to the terms in the forced choice paradigm could lead to post hoc analysis of the stimulus, leading to an evaluation that was not made intuitively (Zajonc 1968). Because it is essentially impossible to determine whether these evaluations are made in the absence of a prompt, most of the research that followed from these findings has been in the marketing and advertising fields rather than in psychology or cognitive science (e.g. Meyers-Levy & Peracchio 1994).

A more recent study asked whether the direction of a character's motion could determine what the audience thought of them and used coded responses to open-ended questions to determine viewers' subjective opinions of a short film. The study claims that viewers freely rated the 'left-to-right' film more positively and the 'right-to-left' film more negatively. However, upon examining the experiment's methods, it becomes evident that the open-ended responses offered very little information about viewers' subjective evaluations of the films. Instead, data on viewer evaluation was pulled from forced-choice semantic differentials (Egizii et al. 2012). Though the method

used in this experiment is less rigid than that of the 1987 study, it has the same problem: you cannot tell whether the evaluations occur in the absence of the question.

Forced-choice ratings and responses to open-ended prompts have offered intriguing preliminary evidence for a relationship between filmmaking decisions and audience experience. Examining attention, an implicitly quantifiable property of the viewing experience, might offer complementary findings to the existing research, given some of the limitations of forced-choice and open-ended methods. Cognitive science, and attention research in particular, has shown many times that our intuitions about our minds can be inaccurate or miss important details altogether. Take the example of inattention blindness: when another task or distraction draws enough of a person's attention, they can completely overlook things that otherwise might be very salient, such as a person in a gorilla suit walking onto a basketball court during a passing drill (Simons & Chabris 1999). Intuition and anecdote are extremely important tools for communication and thought. They allow us to understand complex concepts and share our individual understanding of the way the world works with others. Film theory and past research has proposed a wide range of psychological effects that cinema may have on audiences. Much the same way that cognitive scientists interested in morality use the canon of philosophy as their raw material and inspiration, we can use the intuitions of film theorists and filmmakers as a starting point for the scientific exploration specific film phenomena.

Background

The present study aims to relate forward and backward dolly movements and the spatial distribution of audience attention.

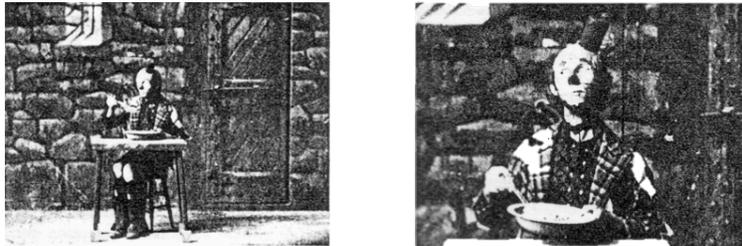


FIGURE 2 – Still frames from the first known dolly move in a professionally produced and distributed film, *Hooligan in Jail* (1903)

Dollying

A *dolly in* is a shot in which the camera moves forward in physical space, getting closer to the actor or object of interest over the course of the shot. Also known as a push-in, the dolly in was among the first few camera movements employed in cinema, first used in 1903 (*Hooligan in Jail*, FIGURE 2). Wheels enabled early cameras, which were extremely heavy, to move for the first time, and the first move ever performed in a widely distributed film was a dolly in. The idea of putting a camera on wheels took off until the early 1920s as part of the studio system, and today, dollies are considered an essential part of most professional film and television sets. Today, dolly moves are so common in TV and film that many audiences do not notice them when they happen. Though some dolly moves are fast enough to catch your attention, a vast majority of

modern push-in shots are not; as a viewer, you may only notice the move if you explicitly look for it. The camera movement in these shots is often below the threshold of conscious motion detection. This kind of dolly move is metaphorically similar to the experience of waiting in line at the DMV. You may be vaguely aware that the line is moving forward, and it is getting closer and closer to your turn, but it is impossible to notice any real progress.

Whether salient or subtle, the visual experience of watching a dolly move is principally the same as passively riding on a moving walkway, like the kind you might find in a large airport. Imagine you are walking through the terminal to your gate, a few minutes to spare before boarding. Miraculously, you find a moving walkway with no one riding else on it. You step onto the walkway and, because it is extremely useful for this analogy, hold your body and head completely still, with your eyes focused straight ahead, down the path of the walkway. When you first step onto the walkway, you can see a speck of blue, hanging from the ceiling at the end of the walkway. A pillar, a few feet away from you on your right, takes up most of the right side of your visual field, with some sort of restaurant visible, but tiny, on the right at the very end of the walkway. A few moments later, you glide past the pillar, revealing Gate A4, which was previously hidden behind it. Rows of chairs sit between you and the gate desk. At that moment, the blue speck at the end of the walkway appears larger; it is apparently a rectangular sign, but you still cannot read it. A few seconds later, and the rows of chairs

are no longer between you and the desk, although you can still see them at the extreme right of your visual field. Wait another few moments, and you can no longer see Gate A4 at all; you are nearing the end of the walkway and the tiny restaurant is now a full-size airport Friday's. What was once a speck of blue is now an enormous sign, the bottom half of which fills the top third of your visual field and warns you 'The moving walkway is ending.' Throughout the ride, your field of view and the relative positions of the objects within the airport stayed the same, but what you were able to see and the relative distances between you and the sign, the Friday's, the pillar, and the desk at Gate A4 changed constantly. Some objects exited the frame as you moved past them, others grew larger as neared them, and still others were blocked or revealed by the objects closest to you. The extremely slow dolly moves mentioned above are more like riding just the last five feet of a moving walkway that has been slowed to one-tenth of its typical rate.

It is important note that a dolly in is different from a *zoom in*: while a dolly in involves physically getting closer to something, during a zoom in, the camera remains static while the internal focal length of the lens increases.^{3,4} Instead of decreasing the

³ Focal length is the distance between the final lens element (most camera lenses contain more than one piece of glass; each piece of glass is called an element) and the sensor of the camera, typically measured in millimeters. The longer the focal length, the narrower the field of view; the shorter the focal length, the wider the field of view (Kingslake 1989). For examples of both a dolly in and a zoom in, and to see the differences between the two, watch https://www.youtube.com/watch?v=r988E_J14FM.

⁴ The zoom in is likewise an intriguing technique to investigate, especially given that the visual experience of zooming in is impossible with the biological image processing hardware that humans

relative distance between audience and object as in a dolly in, a zoom in decreases the audience's field of view. Go back to the beginning of the moving walkway in the airport. You can see the blue warning sign at the end of the walkway as only a speck. If you were holding a camera, and zoomed in on the sign to be able to read it from that vantage point (human eyes cannot physically zoom in any meaningful way, so the camera is necessary for this exercise), you would never see Gate A4, and might only catch the Friday's sign in the bottom corner of the frame. The sign would fill your field of view, but your field of view might only be a few degrees.

Within the film world, educators often assert that a dolly in will guide audiences to attend to focal objects, or, more generally, "direct the audience's attention" (Hilliard 2014; Flacy 2016). However, some theorists make highly specific claims about the effect of a dolly in. Take this example from a detailed taxonomy of camera movements:

However, in its conventionalized form, the push-in is not merely a dramatic intensifier, but is also psychologically suggestive, i.e. aided by a performative or musical cue, the push-in customarily underscores or suggests psychological activity occurring within the mind of the character that the camera is advancing towards.

Nielsen 2007

Though filmmakers have created films with this principle, and others, in mind for decades, there is nothing to back up these assertions other than the words of other

possess. Perhaps the lack of ecological validity could provide complementary results to those obtained from this investigation of the dolly in. I will propose an investigation of this technique as a potential realm for future study.

filmmakers and theorists. To what extent are these claims accurate? How can we scientifically test how a dolly movement influences viewer attention?

Attention

Many factors are known to shape the way we distribute our attention, and these features are usually divided along a dichotomy: endogenous/overt/voluntary factors, and exogenous/covert/involuntary factors. Let's return to the moving walkway at the airport. In a shocking plot twist, you are meeting a friend at the Friday's at the end of the moving walkway for dinner before your flight. You know that your friend is already there, sitting at a table at the restaurant, so as you near the restaurant you actively visually search the seating area for her. This is an example of endogenous or top-down attention: you have actively and voluntarily deployed it to the Friday's seating area. However, not all endogenous factors are voluntary. Working memory can have a top-down effect on attention (imagine trying to remember a phone number while searching for something), and natural patterns of brain activity can affect sustained attention: much like eyes blinking or a webpage that automatically refreshes, attention seems to cycle on and off of maximum performance when it is directed at something for

extended periods of time (Park, Kim & Chun 2007; Van Rullen 2013; Posner 1980; Noudoost et al. 2010).⁵

Back at the airport, you are riding the moving walkway, looking for your friend at Friday's. You know she is wearing a purple shirt. You notice every purple suitcase, every order of (purple) cabbage and eggplant. The Minnesota Vikings are playing a football game on television, and though you know your friend is not on TV, you still look at it. As you glance around the seating area, you are irresistibly drawn to look at one particular patron, a young man with a huge blonde Mohawk, even though your friend is a brunette with curly hair. A person stands up at the table next to Mohawk-man. The video board at the A4 gate desk changes: it goes from a dark grey to bright white. You were focused on finding your friend, but you can't help but notice the board. An instant later, a bird flies in front of you, just inches away from your face. All of these cues, the other purple objects, the blonde Mohawk, the bright video board, are exogenous cues to attention. They are properties of stimuli that fit into two general categories: surface properties and spatiotemporal properties. Surface properties that have been demonstrated to have attentional impact include color (like the purple suitcases, food, and football jerseys), form (the Mohawk), luminance (the white video board), salience (why is there a bird in the airport?) and relevance (Jost et al. 2005;

⁵ Endogenous attention is also described as "goal-oriented," and in the case of object attention, the goal is to keep track of the object and discern information about it for the sake of completing an object-relevant task (Moore 2006). Often, in the lab, the task is simply 'keep track of the object for the duration of the trial' (He, Cavanagh & Intriligator 1997; Intriligator & Cavanagh 2001).

Turatto & Galfano 2000; Fecteau & Munoz 2006). Covert properties can also be more “high-level,” like human faces and text (Theeuwes & Van der Stigchel 2006; Cerf, Frady & Koch 2009).

However, for our study of dolly shots and attention, the most relevant cues for us to examine are spatiotemporal. Several kinds of motion, like the bird flying past your face, or the biological motion of a person getting up from their chair while you are looking for your friend, have been identified as attentional cues (Zhao et al 2014; Pratt et al 2010). Two specific kinds of motion, known as looming and receding motion, have been shown to capture attention extremely well, and may be especially relevant to our study of dolly camera movement and attention. In the most basic sense, *looming motion* is a kind of motion where objects move away from a central vanishing point very quickly; its opposite is *receding motion*, where objects move at speed towards a central vanishing point.

In one of the most striking demonstrations of this effect, researchers showed participants displays of random dot patterns (moving white dots against black backgrounds), which were divided into two groups, left and right hemifields, on the display. Every trial began with all of the dots moving randomly, but after a little over half of a second, dots in one of the two hemifields would start to move in a coherent pattern of looming, receding or unidirectional (up, down, left or right) motion. Soon after the start of coherent motion, a target probe would appear on one side of the

screen or the other. When the probe appeared on the same side of the screen as a field of dots simulating looming motion, participants were significantly faster at detecting them than in any other case. This result proved that looming motion in one part of the visual field has the power to capture and spatially orient attention towards that region (von Mühlenen & Lleras 2007). A continuation of this line of research found that receding motion can also attract attention to specific regions of the visual field, but looming motion does so more strongly and enables more efficient processing of targets inside the field of looming motion (Rossini 2014).

Inward and outward camera movement and looming and receding motion share directional properties of a visual principle known as optic or optical flow: “the change of structured light in [an] image... due to a relative motion between the [optical sensor] and the scene” (Raudies 2013). To put the term to use: looming motion is characterized by optic flow pointing out from the center of the image, and receding motion is characterized by optic flow pointing towards the center of the image.⁶ The vectors of optic flow in a dolly in point the same direction as in looming motion, and a dolly out shares directional optic flow properties with receding motion. However, dolly

⁶ The concept of optic flow is far more nuanced and complex than the simplification presented here. It is of particular interest to both vision scientists and computer scientists attempting to engineer computer vision. Computers tracking optic flow typically represent it through the comparison of displacement vectors of pixels and clusters of high contrast pixels over time (Brox et al. 2004). The prevailing theory on how the brain creates stable cognitive representation of objects and environments undergoing optic flow suggests that orientation detection mechanisms, such as those demonstrated by Gray and Singer during single cell recording (1989), are responsible for the capacity. For more background on the conceptual underpinnings of optic flow, see Koenderink 1986.

movement in and out are not identical to looming motion. In all of the looming motion research that the author is aware of, the simulated optic flow would be the result of a rate of motion much faster than most, if not all, dolly movements. In looming motion studies, the intention is typically to simulate the visual experience of an object flying at your head fast enough to make you duck. Even for the fastest dolly movements, the magnitude of optic flow is nowhere near that size. For this reason, it is unclear whether the optic flow properties of dolly movement would have the same attentional effect as the stimuli used in looming motion studies.

Additionally, looming motion research, and a good chunk of attention research in general, focuses on attentional capture: what features are capable of drawing attention initially. The focus of this study is different and, to the author's knowledge, unique among studies of visual attention. We are interested in whether dolly movement has a continuous effect on the attention of a viewer; if there a general orienting effect towards the middle or the edge of the frame that can be predicted by the kind of camera movement used to create the shot. More fundamentally, can dolly movement, a spatiotemporal property of the stimulus, create an attentional set for the viewer, influencing their noticing behavior throughout the visual field?

The Present Study

The present experiment will use a probe experimental paradigm to investigate the spatial attentional effect of dolly in and dolly out camera movement that is common in cinema. The probe design of the experiment is in the same family as that employed in the first looming motion study, described above, motion, but has been modified in particular from a task used in Alvarez and Scholl's 2005 study. To test the hypothesis that camera movement spatially orients attention towards central or peripheral regions we will examine the accuracy of probe detection of participants while watching movie clips, analyzing their response across three eccentricities radiating from the center of the image.

If we use the existing data on looming and receding motion as analogy to the optic flow of dolly movement, dollying in should "direct the audience's attention" in some way, due to the outward optical flow of the image and the ability of looming motion to capture attention, and dollying out might have a similar, though less extreme, effect (Skarratt, Cole & Gellatly 2009; Rossini 2014). However, none of the existing research has yet examined the nature of attention allocation within a region undergoing optical flow; instead, they have all compared the attentional impact of regions undergoing optical flow next to regions undergoing random motion (Franconeri & Simons 2003; Judd et al. 2004). Perhaps the optic flow of dollying in and

dollyng out, respectively, will both direct attention centrally, as Rossini insinuated by placing probe stimuli at the center of the optic flowfield (2014).

However, there may not be any systematic effect of the direction of camera movement. Attention and eye fixation might instead be driven by other spatiotemporal or surface properties of the stimulus. For example, human faces or biological motion might exogenously attract viewer attention. Although our initial analysis will examine the relationship between eccentricity and probe detection accuracy, our experimental design allows us to test alternative hypotheses by generating the display locations of the attentional probes randomly, rather than with equal distribution across eccentricities.⁷ Though such analysis will not yet be completed at the time of the submission of this paper, it will continue over the coming months.

Methods

Participants

Twenty-five naïve observers from the Yale/New Haven community (all with normal or corrected-to-normal acuity) participated in exchange for small monetary compensation. The number of participants was based on common sample sizes in

⁷ We initially considered the possibility of overtly distributing probes equally across the three eccentricities, given that the central eccentricity has a substantially smaller area than the most peripheral. However, the ability to analyze data for other trends and the value of removing potential causal links between the design and results of the experiment outweighed the experimenters' desires for neat data.

similar vision science experiments and was defined before the experiment started. A programming hiccup limited the first five participants to an average of 41.2 trials (S.D. = 1.48) of the 55 intended trials.

Stimuli

Stimuli were composed of 21 experimenter-created clips and thirteen filler clips from professionally produced and distributed films. Three versions, dollying in, dollying out, and static, of 7 scenes were created using a Sony NXCAM FS700 and Odyssey 7Q external recorder, with Sony Vario-Sonnar T* DT 16-80mm and Rokinon 24mm Cine lenses, mounted on a Matthews Doorway Dolly with pneumatic tires. Video was captured to the Odyssey recorder in 4K Apple ProRes 422 HQ and initial editing was performed in Final Cut Pro X. The files were stabilized to correct for unsteadiness or veering and letterboxed (black bars above and below the image) to an aspect ratio of 16:9 in Adobe AfterEffects, then exported at a resolution of 1024x768 pixels and compressed into an MPEG-4 codec in Compressor. The clips had an average duration of 26.4 seconds ($n = 7$, S.D. = 4.9). See Appendix A for representative still frames from each created scene.



FIGURE 3 – Diagram of Eccentricities

These stimuli were created specifically for this experiment with two goals in mind. First, on a statistical level, matched triplets of dolly in, dolly out and static shots of the same composition are extremely useful, as they offer a second frame of comparison beyond category of camera movement. Second, they were created to look and feel as though they were taken from scenes in professionally produced movies. The content and location of the scenes varies. They include a woman looking out a window pensively, a man standing in a hallway and looking menacingly towards the camera, groups of people in conversation, and a door obstructed by piles of junk, among others. The intent was for viewers to have a sense that these shots possessed some narrative or emotional significance, but had been removed from the context that lent them that meaning.

The clips from existing films were extracted from movies produced between 1914 and 1945 ($n = 4$, mean = 1932, S.D. = 14) and varied in duration from 10 to 34 seconds ($n = 13$ mean = 18.9, S.D. = 7.1).

Procedure

Participants performed a probe detection task, and underwent 55 trials. Subjects viewed the film clips unrestricted (estimated 65cm from the monitor) and without sound on a Dell CRT computer monitor with a resolution of 1024 x 768 pixels and a refresh rate of 60Hz. The clips were presented using a custom MATLAB program.

Participants were instructed that as they watched the clip, “small white dot(s)” would appear intermittently throughout the clips and that their “job is to press the ‘space’ key as soon as possible whenever [they] see the dot appear.”

The probes were white disks with a diameter of 0.13° visual angle presented at a random location, regardless of the content of the image, meaning that some appeared on top of objects, others on the background, and everywhere in between. The probes were superimposed on the movies with a margin of 30 pixels (0.95° visual angle), for roughly 200 ms. The first probe appeared roughly 1-4 seconds after the onset of the clip. The SOA between probes was between roughly 2.5 and 5 seconds. A timed-out miss was recorded 1250 ms after the offset of the probe. A false alarm was recorded if the participant pressed the spacebar before first probe, within 50 ms after the onset of a probe, and after 1250 ms after the offset of a probe and before the onset of the next. Otherwise a hit is recorded.

Results

Three eccentricities were defined to compare probe detection accuracies: center, middle and far. All three were centered on the middle of the display, with the center eccentricity defined as a circle with a diameter of half of the height of the probe region. The middle eccentricity was a circle with a diameter of the full height of the

probe region, minus the area of the center eccentricity. The far eccentricity was composed of the remaining area of the probe region (FIGURE 3).

As depicted in FIGURE 4 and TABLE 1, our initial parametric analysis of probe detection accuracy across eccentricities indicates that there is not a relationship between camera movement and eccentricity to probe detection percentage. Probe detection was entered into a two-way repeated-measure ANOVA. A significant main effect was found for eccentricity on detection accuracy [$F(2, 48)=47.09, p<.001$], but there was no main effect of camera movement [$F(2, 48)=0.25, p>.250$]. No interaction was found between the two main effects [$F(4, 96)=0.95, p>.250$]. Although parametric analysis seems to dismiss the effects of camera movement, exploratory non-parametric analysis of within-subject probe detection accuracies by movement type suggests that there may yet be an effect present. Subjects performed better at probe detection in the middle eccentricities on dolly in shots than on dolly out shots [Binomial test: $B(25, 7, p<.05)$] and had worse probe detection at the furthest eccentricity on dolly out shots when compared to static shots [Binomial test: $B(25, 7, p<.05)$].

Though these two tests would not survive multiple comparisons, they do suggest a glimmer of hope that there may be some effect in play, but the limitations of the existing dataset and the analysis so far do not support such a claim well. We also explored whether there were effects of camera movement within particular movies, and all had $p>.05$ except for one: center + middle probe detection rate was better for most

subjects on the dolly in version of the scene called *outside* than on the static version

[Binomial test: $B(23, 7, p < .1)$].

FIGURE 4 – Probe Detection Rates by Eccentricity and Camera Movement Type

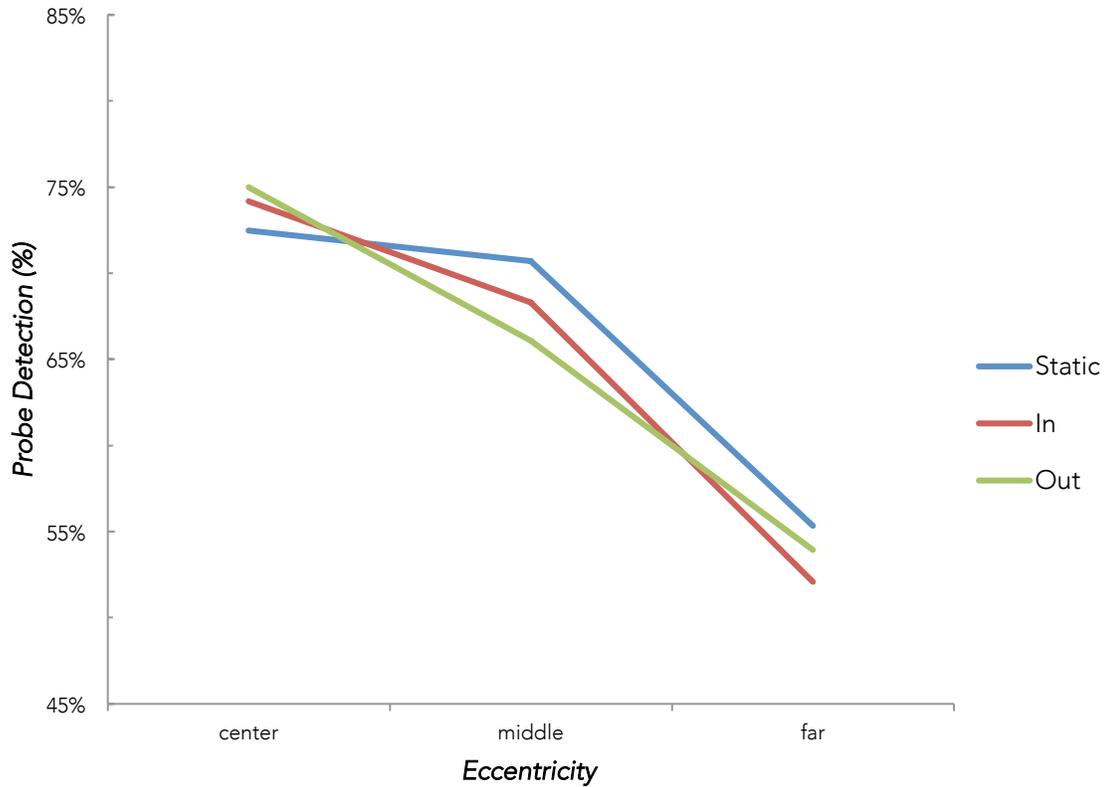


TABLE 1 – Two-Way Repeated Measure ANOVA of Probe Detection Rate against Camera Movement and Eccentricity

Source	SS	df	MS	F	p	pEtaSq
Dolly	0.008	2	0.004	0.247	0.782	0.01
Ecc.	1.617	2	0.808	47.088	0.000	0.662
Subject	2.877	24				
D*E	0.041	4	0.01	0.951	0.438	0.038
S*D	0.749	48	0.016			
S*E	0.824	48	0.017			
S*D*E	1.022	96	0.011			

Discussion

We set out to answer the question of whether dolly movement inward or outward had an effect on where viewers attend within the frame of a film. This experiment and the present analysis specifically looked for an effect of the direction of camera movement on probe detection accuracy across three eccentricities and did not identify any such effect. Exploratory analysis suggests that continued investigation, at least within the existing dataset, may be worthwhile, but it is impossible to conclude anything more at present.

Surveying a path forward for this research question and others like it, we will discuss the limitations of the present experiment and analysis, other methods that could be applied to this question, a broader research program for the study of attention and film, and future directions for filmmakers, scientists and the viewing public.

Limitations

Some of the trends described in the **Results** section above highlight specific limitations of the present design and analysis. After addressing these trends, we will examine some of the more general limitations of this study. The most interesting but perhaps most confusing of these patterns is the probe detection advantage in the middle eccentricity for dolly in shots over dolly out shots. None of our hypotheses,

derived from either film theory or vision research, predicted a probe detection advantage in the intermediate eccentricity for any condition. One possibility is that due to the small number of trials at the central eccentricity, we have not yet been able to identify a trend in this region above the data's noise floor. Continuing analysis or simply more trials might find an even greater advantage in the central eccentricity for dolly in over dolly out shots, suggesting a gradient of attentional deployment. This interpretation is purely speculation at this point, as there is very little complementary evidence for this trend.

Another intriguing element of the current data analysis is *outside's* impact on probe detection accuracy by eccentricity. What features of this scene created a detectable advantage of central + middle over peripheral targets, where the other shots did not? Notable distinctions that set *outside* apart from the other movies include: it is the fastest dolly move of the group, it is a very bright scene, the apparent focal objects (three people) are located at the center of the screen, and there is substantial biological motion from other people entering and leaving shot in the background. It is impossible to say which, if any, of these factors played a role in establishing trend with the current analysis, but examining detection rates with a finer-toothed comb, perhaps by comparing probe detection rates across coded regions of interest in the scene, or another form of statistical investigation, might lend further insight, as might an exploration of probe reaction time.

More generally, this experiment serves as an introductory exploration of the study of film technique using cognitive science methods, and as such, has many flaws that have influenced the quality of data collected and deserve to be addressed in future studies. As mentioned above, the limited number of probes shown in the central eccentricity led us to combine the central and middle regions for binomial analysis. Longer clips that permit more probes per session, as well as more created clips in general would provide a larger sample of probes per subject. Another critique of this design is that although the stimuli are ecologically valid, they are highly uncontrolled. Plainly, there is so much going on in each clip that the main effect of camera movement may have been clouded by the variety of other factors that can capture attention. The noise floor for our data is extremely high. However, because all of the stimuli are equally uncontrolled, the data for each stimulus is equally noisy, with camera movement as the independent variable across all stimuli. However, with so few stimuli and so few probes per subject, this noise may have covered up trends in our data.

Additionally, many of the subjects recognized actors and locations in the created clips, as they were filmed around campus using students as actors. A vast majority of participants were Yale undergraduates, and every single participant recognized at least one location or actor. This may have added to the noise of the experiment as subjects may have disengaged from the task at hand to think about whether they recognized the actors and locations in the films. In the future, stimuli should be created in locations

unfamiliar to participants using actors that they neither know personally nor recognize from other films.

Finally, the probe task itself may have put subjects in a different attentional state than is typical for a person who is watching a movie. Instead of paying attention to the action and content of the movie clips themselves, they may have disengaged from the films and aimed for the highest possible probe detection score by using a strategy totally unlike their typical film or television viewing strategy. It is possible that a main effect of camera movement only appears when viewers have a context- and task-relevant attentional set.

Other methods?

Other experimental methods, such as eye tracking and search tasks, could be applied to the study of camera movement and attention. Eye tracking, in particular, has a number of strengths that would complement the current probe paradigm. Most prominently, eye tracking would allow subjects to view clips in a more typical attentional state, removing the task demands of the probe design. It would also offer a nearly continuous picture of each subject's fixations throughout the clips: where the probe design has a sample rate of roughly one probe every 2.5s, eye trackers typically collect at least 50 samples every second. Though fixation location is not a direct

representation of covert attention, together with a probe paradigm or another similarly capable method, it can provide a more complete picture of the viewing experience.

A broader research program on cinematic attention?

Using the methods outlined in the report and suggested in the section above, a broader research program on cinematic attention could be designed around the formalist distinctions of the seven types of camera movement. These movements could be tested with methods derived from those used in this experiment and stimuli created specifically to test those movements. Zooming and Dolly Zooming could be studied using essentially the same procedure as was used in the experiment, given that they share an axis of apparent motion into or out of the screen. The visual experience of a zoom in is a narrowing angle of view over time, showing a smaller region of the scene in an image of the same size. It occurs when the distance between the final lens element and the image sensor increases over time. A zoom out is the opposite: a wider angle of view over time, revealing a larger region of the scene in an image of the same size. A dolly zoom is also known as a Vertigo Zoom (it was used to great effect in Alfred Hitchcock's *Vertigo*, and is often described as simulating the feeling of vertigo), and involves dolly back while zooming in, or zooming out while dolly in. A Dolly

Zoom has an unusual visual effect of keeping the focal object the same scale within the frame, while 'compressing' or 'decompressing' the background.⁸

Simple pans right and left could also be examined with a probe paradigm. A pan looks much the same as the visual experience of turning your head. When panning, the camera sits at a fixed point and rotates about an axis that is typically a few inches behind the sensor. Given film theory and anecdote, and the fact that pans move laterally, rather than in or out, it may prove more useful to compare probe detection rates in the leading, center and trailing thirds of the frame, rather than using eccentricity as a dimension of interest. The very same design could be used to test the attentional impact of trucking: when the camera moves from side to side. A trucking shot is very similar to the visual experience of looking out of the window of a moving car. A comparison between the attentional effects of trucking and panning could provide intriguing parallels to the comparison of dollying and zooming. Both pairs are characterized by matching patterns of optic flow, but markedly different physical movements of the camera.

⁸ Neither a zoom nor dolly zoom is an ecologically valid stimulus: it is impossible to experience a Zoom or Dolly Zoom without an external optical device. It is difficult to say whether this fact will play any role in their attentional impacts, but they might add another dimension of comparison for pure dolly shots, given that all three have radial optic flow patterns.

Future directions

The limiting factor for scientific research on filmmaking practice and viewing experience has been the lack of cross-talk between filmmakers and scientists. In order to examine these questions, cognitive scientists need stimuli that are created specifically with research questions in mind, but that still look like movies.

Though the current study does not offer definitive conclusions regarding the relationship between camera movement and the spatial distribution of attention, it does offer a foundation upon which to build a more robust research program to examine this question, and the cognitive science of filmmaking practice and viewing experience more broadly. As video content becomes environmentally pervasive (e.g. animated billboards, videos in Facebook newsfeed that play whether you want them to or not, the prevalence of YouTube, smartphones, refrigerators with screens on them, etc.) it is increasingly important to curate and develop an understanding of what draws audiences to attend to one thing or another within a video, and of how the perception of movie/videos differs from naturalistic perceptual experience. The experimenters strive to obtain more complete conclusions with the continuation of their collaboration, combining shared interests in the experimental exploration of filmmaking practice, aesthetics and viewing experience with disparate expertise as perception researcher and filmmaker.

References

- Alvarez, G. A., & Scholl, B. J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General*, 134(4), 461.
- Alwitt, L. F., Anderson, D. R., Lorch, E. P., & Levin, S. R. (1980). Preschool children's visual attention to attributes of television. *Human Communication Research*, 7(1), 52-67.
- Barten, P. G. (1999). *Contrast sensitivity of the human eye and its effects on image quality* (Vol. 72). SPIE press.
- Bordwell, D. (1997). *On the history of film style*. Cambridge, MA: Harvard UP.
- Brox, T., Bruhn, A., Papenber, N., & Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *Computer Vision-ECCV 2004* (pp. 25-36). Springer Berlin Heidelberg.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, 51(13), 1484-1525.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of vision*, 9(12), 10-10.
- Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, 49(6), 1407-1409.
- Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta psychologica*, 149, 69-77.
- Cutting, J. E. (2015). The framing of characters in popular movies. *Art & Perception*, 3(2), 191-212.
- Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, 2(6), 569-576.
- Cutting, J., & Iricinschi, C. (2015). Re-Presentations of Space in Hollywood Movies: An Event-Indexing Analysis. *Cognitive science*, 39(2), 434-456.
- DeLong, J. E., Brunick, K. L., & Cutting, J. E. (2012). Film through the human visual system: finding patterns and limits. *Social science of cinema*. 123-137.
- Egizii, M. L., Denny, J., Neuendorf, K. A., Skalski, P. D., & Campbell, R. (2012). Which Way Did He Go? Directionality of Film Character and Camera Movement and Subsequent Spectator Interpretation. *ResearchGate*.
- Eisenstein, S., & Leyda, J. (1975). *The Film Sense*. Rev. ed.
- Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., Kuzmova & Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 503-514.
- Ewert, J. P. (1971). Single unit response of the Toad's (*Bufo americanus*) caudal thalamus to visual objects. *Journal of Physiology*, 74, 81-102.
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences*, 10(8), 382-390.
- Flacy, M. (2016). 4 Uses of a Camera Dolly in Movies. *Steve's Digicams*
- Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65 (7), 999-1010.
- Giannetti, L. (2011). *Understanding movies* (12th ed.). Boston, MA: Allyn & Bacon.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86(5), 1698-1702.
- Greenlee, M. W. (2000). Human cortical areas underlying the perception of optic flow: brain imaging studies. *International review of neurobiology*, 269-292.

- He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in cognitive sciences*, 1(3), 115-121.
- Hilliard, R. (2014). *Writing for television, radio, and new media*. Cengage Learning.
- Hooligan in Jail* [Motion picture]. (1903).
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive psychology*, 43(3), 171-216.
- James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.
- Jost, T., Ouerhani, N., Von Wartburg, R., Müri, R., & Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, 100(1), 107-123.
- Jouen, F. (1990). Early visual-vestibular interactions and postural development. In H. Bloch & B. I. Bertenthal (Eds.), *Sensory-motor organizations and development in infancy and early childhood* (pp. 199-215). Dordrecht: Kluwer.
- Jouen, F., Lepecq, J.-C., Gapenne, O., & Bertenthal, B. I. (2000). Optic flow sensitivity in neonates. *Infant Behavior & Development*, 23, 271-284.
- Judd, A., Sim, J., Cho, J., Muhlén, A., & Lleras, A. (2004). Motion perception, awareness and attention effects with looming motion. *Journal of Vision*, 4(8), 608-608.
- Koenderink, J. J. (1986). Optic flow. *Vision research*, 26(1), 161-179.
- Kraft, R. N. (1987). The influence of camera angle on comprehension and retention of pictorial events. *Memory & cognition*, 15(4), 291-307.
- Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in cognitive sciences*, 7(1), 12-18.
- Moore, T. (2006). The neurobiology of visual attention: finding sources. *Current opinion in neurobiology*, 16(2), 159-165.
- Motion Pictures. (2012). *The Thomas Edison Papers*. Rutgers University. Retrieved from <http://edison.rutgers.edu/pictures.htm>
- Nielsen, J. I. (2007). *Camera Movement in Narrative Cinema*. Ph.D. Dissertation: University of Aarhus.
- Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current opinion in neurobiology*, 20(2), 183-190.
- Park, S., Kim, M. S., & Chun, M. M. (2007). Concurrent working memory load can facilitate selective attention: evidence for specialized load. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1062.
- Peracchio, L. A., & Meyers-Levy, J. (1994). How ambiguous cropped objects in ad photos can affect product evaluations. *Journal of Consumer Research*, 190-204.
- Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25.
- Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive! Animate motion captures visual attention. *Psychological Science*, 21(11), 1724-1730.
- Raudies, F. (2013). Optic flow. *Scholarpedia*, 8(7), 30724.
- Rauschenberger, R. (2003). Attentional capture by auto-and allo-cues. *Psychonomic Bulletin & Review*, 10(4), 814-842.
- Regan, D., & Beverley, K. I. (1978). Looming detectors in the human visual pathway. *Vision research*, 18(4), 415-421.
- Rind, F. C., & Simmons, P. J. (1997). Signaling of object approach by the DCMD neuron of the locust. *Journal of Neurophysiology*, 77, 1029-1033.

- Rossini, J. C. (2014). Looming motion and visual attention. *Psychology & Neuroscience, 7*(3), 425.
- Schmuckler, M. A., Collimore, L. M., & Dannemiller, J. L. (2007). Infants' reactions to object collision on hit and miss trajectories. *Infancy, 12*, 105-118.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception, 28*(9), 1059-1074.
- Skarratt, P. A., Cole, G. G., & Gellatly, A. R. (2009). Prioritization of looming and receding objects: Equal slopes, different intercepts. *Attention, Perception, & Psychophysics, 71*(4), 964-970.
- Skarratt, P. A., Gellatly, A. R., Cole, G. G., Pilling, M., & Hulleman, J. (2014). Looming motion primes the visuomotor system. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 566.
- Styles, E. A. (2005). *Attention, perception and memory: an integrated introduction*. Psychology Press.
- Sun, H., & Frost, B. J. (1998). Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nature Neuroscience, 1*, 296-303.
- Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62*, 626-641.
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta psychologica, 135*(2), 77-99.
- Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657-665.
- Thompson, K. (1999). Storytelling in the new Hollywood. *Understanding Classical Narrative Technique*. Cambridge, MA: Harvard UP.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*(4), 410-433.
- Turatto, M., & Galfano, G. (2000). Color, form and luminance capture attention in visual search. *Vision research, 40*(13), 1639-1643.
- VanRullen, R. (2013). Visual attention: A rhythmic process?. *Current Biology, 23*(24), R1110-R1112.
- von Mühlenen, A., & Lleras, A. (2007). No-onset looming motion guides spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1297.
- Weir, P. T., Schnell, B., & Dickinson, M. H. (2014). Central complex neurons exhibit behaviorally gated responses to visual motion in *Drosophila*. *Journal of neurophysiology, 111*(1), 62-71.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology, 9*(2p2), 1.
- Zhao, J., Wang, L., Wang, Y., Weng, X., Li, S., & Jiang, Y. (2014). Developmental tuning of reflexive attentional effect to biological motion cues. *Scientific reports, 4*.

Appendix A – Stills from Experimenter-Created Stimuli

