



Considering Eye-tracking as a Validation Tool in Cinema Research

Giorgos Dimitriadis

Aristotle University (Thessaloniki, Greece)

E-mail: geodim@enl.auth.gr

Abstract. The use of eye-tracking in data collection, when accompanied by the proper research questions and methodology, is a powerful tool that may provide invaluable insights into the way viewers perceive and experience movies. Film theory can use eye-tracking to test and verify research hypotheses not only with unprecedented accuracy, but also with the ability to address a significant variety of theoretical questions. Eye-tracking can help build contemporary film theory by supporting its various fields of research, and also even assist the production of films themselves by helping filmmakers make more informed creative decisions. The present article is an overview of eye-tracking and its gradual implementation in cinema research; in the context of discussing some recent examples of academic work based on eye-tracking, it considers the technology of eye-trackers and the way in which human vision handles visual information on screen. By testing the attentional behaviour of viewers, eye-tracking can produce more solid answers to questions regarding the way films are experienced; therefore, it may very well prove to be the spearhead of a more robust body of film theory in the near future.

Keywords: eye-tracking, gaze data, visual perception, cinema, cognitive film theory.

Introduction

Cinema has always been a showcase for new imaging technologies. Driven by the power of a constantly evolving digitization, film production is perpetually on the verge of something new, a condition which more often than not produces corresponding dynamics for the research approaches that serve it. Probably among the most significant of these approaches, currently under the spotlight of a growing and, most notably, interdisciplinary body of researchers, is the use of eye-tracking. Eye-tracking is expanding rapidly to all aspects of research

that entail human vision, such as psychology of perception, marketing and commerce strategies, interactivity in software and web design, education, hardware operation, etc. Cinema was bound to be included in the list, especially since cognitive approaches entered film theory. The continuous technological development of eye-tracking systems has made their use significantly easier and far less invasive, also leading to a remarkable improvement in their accuracy and overall usability with a much broader body of participant groups. This, in turn, opens up tremendous possibilities in cinema research, given the extreme popularity of the medium among a considerable number of target groups and their combinations as well as its inherent diversity in style and technique. The same diversity should naturally be expected in cinema-related research that is based on eye-tracking as well.

The tools that eye-trackers have made available to researchers with the development and refinement of relevant technology over the past few years have steered research towards relatively new directions. Neuroscientific testing and data collection are encrusted with an aura of authority that provides them – and, therefore, the conclusions drawn from their analysis as well – with a desired impact of scientific credibility. Regardless of the eventual validity of each of these results, which is still subject to correct methodology related to a number of parameters ranging from the overall design of the study to the interpretation of results, such approaches are difficult to be matched by what could potentially be a passion-driven, but not necessarily scientifically backed film theory.

Eye-tracking can work equally well with currently emerging theories as well as retrospectively. Tim Smith makes a solid case for the use of eye-tracking in cinema research, noting that the desire to know where exactly people look when watching a movie has only recently been fulfilled after decades of significant efforts, which, nevertheless, have been intuitive and predictive rather than subject to empirical testing and verification (2013, 169). For example, Smith refers to Sergei Eisenstein's beliefs about the possibility of a director guiding the viewer's gaze on screen by carefully considering the way shots are composed; or, to Edward Dmytryk's guidelines on optimal cutting, which he formulated – based on the time needed for viewers to shift their gaze – by measuring his own saccades years before eye-trackers became commercially accessible (2013, 169). Smith notes that cognitive approaches to film theory, in particular, have addressed various issues concerning the mental processes of film viewers as well as the possibility of directorial decisions being able to affect the experience of watching a movie as a whole. He, thus, proposes a framework by which the concerns of cognitive

film theory can be made to communicate with those of empirical psychology and also benefit from its tools and methods, with eye-tracking conveniently offering “a real-time measure of how a viewer is watching and processing a film” (2013, 165–166), a once unattainable luxury for cinema scholars. In addition to this, eye-tracking brings the power of statistical analysis into play: as Adrian G. Dyer and Sarah Pink note, eye-tracking can enable testing a specific visual scenario in a demographically controlled sample of participants (2015).¹ Testing pre-determined hypotheses about the way the viewer’s eyes respond to specific parts of a visual stimulus can enhance even further the findings and conclusions drawn from research based on visual gaze data.

The present article aims at providing an overview and defence of eye-tracking as a useful tool in contemporary research in film, considering the physiology of human vision. Eye-tracking can significantly support film theory and even provide it with entirely new directions. Using this as a starting point, the article first introduces some elementary features and concepts of eye-tracking technology that are particularly pertinent to studies concerning the moving image, before moving on to an essential understanding of the way human vision picks up information from the environment. Within this framework, an overview of some key findings from eye-tracking research on the perception of film will be discussed; these findings reveal the importance of understanding the operations of the human visual system when it encounters a dynamic scene, particularly in mainstream narrative cinema, and, by extension, highlight some of the ways in which eye-tracking can provide more solid foundations to film theory in general.

Understanding the Basics of Eye-tracking

Eye-tracking tracks and records the human gaze, providing patterns of the way human eyes see and observe a practically unlimited array of visual stimuli: if something can be placed in front of a participant’s eyes, it can probably be tracked. Eye-trackers commonly use infrared cameras to capture the reflection of an infrared light source on the eyes, thus being able to map their mobility patterns typically – but not exclusively – on the surface of a screen. According to Andrew Duchowski, eye movement analysis that makes use of systems based

1 The authors stress, though, that eye-tracking can only tell researchers what participants are looking at. It cannot by itself explain “why, what they are experiencing, what their affective states are, nor how their actions are shaped by the wider social, material, sensory and atmospheric environments of which they are part,” for which they suggest turning to phenomenological anthropology (Dyer and Pink 2015).

on gaze detection requires focusing on the way the gaze repositions itself in the visual array with quick jumps (saccades), tracks a moving target (smooth pursuits) and stabilizes itself on specific stationary locations such as objects (fixations), under the assumption that these movements “provide evidence of voluntary, overt visual attention,” without excluding at the same time their involuntary use or nonuse (2007, 47).² By doing so, such systems are capable of recording data about the visual behaviour of participants: researchers can locate the areas and points on a screen that attracted the participants’ gaze and are therefore potentially important as well as track the routes that eyes take as they jump from one fixation to the other. Eye-trackers record and measure the participants’ patterns of looking at something in order to answer the questions “where,” “how long,” and “where next.” The analysis of data collected during an eye-tracking experiment can potentially reveal relationships linking fixations and saccades to cognitive processes, therefore aspiring to offer insights into the foundations on which human perceptual mechanisms operate.

Although eye movement has attracted researchers’ attention for a remarkably long time, the past decades have seen a rapid development in both its methods and technology of analysis. Before that, the techniques used in eye movement research were far from optimal due to the invasiveness of the equipment used, which made the process uncomfortable for participants; early research on eye movement even involved direct contact of the equipment with the cornea (Jacob and Karn 2003, 574; Smith 2013, 166). Technological developments changed that situation by enabling the widespread usability of eye-trackers, thus also enhancing the validity, accuracy and ease of data collection. Eye-trackers using emission of infrared light and one or more infrared cameras capturing its reflection on the retina are the most common ones. A very basic hardware distinction is between screen-based eye-trackers and glasses. Screen-based eye-trackers may be stable lab equipment, in the sense that the entire eye-tracking system is incorporated in the hardware of a screen connected to a computer, which delivers to the participants pre-determined visual stimuli; or they may

2 In relation to this, Keith Rayner lists four types of eye movement: *saccade*, *pursuit*, *vergence* and *vestibular*, stressing the prominence of saccades in the context of eye movement research, as they are “more relevant in typical information processing tasks.” To these he adds three additional types of small movements, i.e. *nystagmus*, *drifts* and *microsaccades*, which are usually treated as “noise” and not considered during eye movement experiments (1998, 373–374). Their existence, though, betrays a “less-than-perfect control of the oculomotor system by the nervous system;” the constantly occurring tremor of *nystagmus*, for instance, renders fixations as “something of a misnomer,” as the word is used to denote stillness of the gaze, whereas in reality eyes never remain completely still (1998, 373–374).

be ultra-portable rods that are connected normally to a laptop and utilize its camera, thus taking advantage of the relative portability of laptops, which makes them practical for use in field experiments and data collection. Glasses, on the other hand, are normally intended for recording real life scenarios (e.g. driving, shopping habits in a store, etc.) outside the laboratory. Infrared trackers as well as cameras are built in the frame of the glasses; trackers are faced inwards so that they can capture the infrared reflection from retinas, while cameras record the participants' visual array as they interact physically with their environment within the given scenario so that fixations and saccades can be mapped on the exact parts of the environment that participants observe. This variation of equipment broadens the range of studies that eye-trackers can accommodate. In addition to this variation, relevant software can further customize the process of data collection: "areas of interest" (AOIs), i.e. specific key areas of the image that are potentially interesting or important for the purposes of a particular research project, are marked and isolated both spatially and, in the case of videos, temporally so that gaze data from those areas only will be considered during data analysis. Isolating AOIs enables the formulation of more targeted research questions as well as more accurate findings. As Dyer and Pink note, such an ability may allow more detailed or specialized studies within film theory, e.g. on the factors that influence the perception of narrative (2015).

The gaze data that can be obtained from an eye-tracker are not only numerical, i.e. sets of data subject to statistical analysis, but also visual, which facilitates a quick interpretation of viewer behaviour. The main forms of gaze data visualization today are "gaze plots" or "scanpaths," "gaze videos," "bee swarms," "heat maps" or "dynamic heat maps," and "focus maps" (Bojko 2013, 218). The eye-tracker software normally superimposes these visualizations on the stimulus (image video, text, etc.) used in the experiment. Gaze plots or scanpaths are sets of circles connected by lines; the circles appear exactly on the fixation positions on the screen, and the lines connecting them show the direction from one fixation to the other. Each set of circles and lines may be coloured to represent different participants and, normally, the longer the fixation is, the bigger the equivalent circle appears. Gaze videos use the same type of circles and lines, but these are presented in motion for each participant, effectively replaying on screen the gaze pattern of the participant exactly as it was recorded. Bee swarms do not show the directions of saccades, they use only small circles or dots to represent fixations, also presented in replay; each circle grows in size during the replay to represent a longer duration of a fixation. The moving gaze points of each participant are

eventually visualized in the form of small moving circles which look like insects flying around in a swarm in front of the screen. Similarly, heat maps also take into account only fixations; the areas of fixation on the screen are covered by a form of coloured clouds, the colour of which reflects the duration of the fixation. Instead of changing size, the level of attention that each area attracted is represented accumulatively with the use of shades of green, yellow and red. The difference between usual heat maps and dynamic heat maps is that the latter offer a live preview of the clouds changing shape and colour during the participants' gaze duration, similar to gaze videos and bee swarms. Finally, focus maps work more or less on the same principles as heat maps but, instead of colours, different levels of clarity are used; i.e. the entire image appears blurry, with only the fixated areas appearing clear and in focus. These varied visualization options may be used to address different sets of data, depending on the aims of the research.

It is obvious that the ability to record and see exactly where, when, and how people look at the environment or a specifically designed visual stimulus provides researchers with a significant level of power to address questions of visual perception and cognition. The data that can be obtained from fixations on specific parts of the visual array as well as the order in which they occur, the direction of the gaze between consecutive fixations, durations of both fixations and saccades, etc. can potentially open a window into the perceptual mechanisms that guide the movement of the eyes and enable humans to acquire the information they need in order to interact with their environment as well as with moving images such as film. The value of such data lies in the fact that these can be derived from actual interaction with the environment, as is the case with eye-tracking glasses, as well as from custom stimuli created to serve tailored experiments, which, in turn, enables the study of actual human observation habits in the context of specific research objectives.

Visual Tasks: How the Viewer's Gaze Works

In order to fully appreciate how eye-tracking can help film theory, and cinema research in general, it is important to have first an elementary understanding of the way vision works, especially when it handles visual stimuli containing motion and rapid changes, such as edited films. The reason for this is the fact that the physiology of human vision poses significant constraints regarding the amount of information accessible to the viewer through the eyes, a fact which should obviously be taken into account when considering visual art and its reception. In addition

to the way vision works, it is also important to understand its connections with mental activity. In fact, Jacob and Karn note that relevant research on the relation between eye fixations and cognitive functions started only after the 1970s, with the development of eye-tracking systems (2003, 575). This section presents some elements that explain the selectivity of human vision as well as a short literature review on the correlation between eye movement and cognitive operations. It is crucial to stress here that most of these works concern a number of different visual scenarios, e.g. reading, abstract stimuli, static media, or dynamic but unedited media such as real-world scenes, rather than edited moving image stimuli such as narrative films; these works indeed support the potential for selectivity that vision has, but the way this selectivity works appears to be different in edited movies. The next section will subsequently treat all this information from the perspective of narrative cinema specifically, so that it becomes clear why narrative film theory probably requires its own branch of eye-tracking research.

The function of the human eyes does not allow either continuous or homogeneous access to all parts of a scene. At the centre of the retina, the fovea is essentially its only part that can see in great detail and focus on whatever it is directed towards, but the visual field that is projected on it is only approximately 2° wide; on the other hand, the peripheral retina is equipped with photoreceptors that are quite sensitive to low light and motion, without the same acuity as the fovea, but covering a much broader area both horizontally and vertically. Therefore, visual resolution deteriorates as we move farther away from the fovea and its maximum acuity, towards the low-resolution but sensitive surface of the peripheral retina (Henderson 2003, 498; Findlay 2004, 136; Hoffman 1998, 119; Majaranta and Bulling 2014, 41; Smith 2013, 167–168; Dyer and Pink 2015; Cohen 2017, 271). As a result, while looking at a scene, the eyes perform a sequence of saccades during which a number of fixated points of focus are collected and processed; these fixations last about a quarter of a second (Hoffman and Subramaniam 1995, 787; Hoffman 1998, 120; Hyönä 2010, 173; Majaranta and Bulling 2014, 40). In fact, since information is only collected during fixations, viewers are effectively blind for the small time that saccades last (Hoffman 1998, 120), an effect called “saccadic suppression” (Matin 1974; Smith 2013, 168). Given the already existing physical human inability to see the entire 360° of visual environment without using head and body movement (Hoffman 1998, 119), the fact that saccades allow only a rather small clear sampling of the environment adds even more limitations to human vision: the part of the scene that is seen clearly and in focus is considerably smaller than the entire scene, a

fact which makes perception “inherently selective” (Cohen 2017, 271). Therefore, the eyes move around a scene and focus on parts of it in a very economically selective way, collecting and combining visual information in the process, thus compensating for their lack of homogeneity in visual acuity. Indeed, the eyes never see a scene clearly in its entirety, instead demonstrating what seems to be a selection path while they fixate successively from one point to another.

Evidence of selectivity is present even in early studies on visual perception – using, in fact, equipment for monitoring eye gaze that would now seem primitive – as well as in more contemporary references in the relevant literature. Only a small fragment of these publications will be mentioned here. One of the cases of research on visual perception that is most cited by eye-tracking scholars is that of Alfred Yarbus, due to the fact that in the 1960s he explicitly associated the direction of saccades to an internal motivation to seek the potentially most informative areas in a viewed scene. Yarbus demonstrated that “the pattern of eye fixations that a given observer produces is influenced by properties of the scene as well as the goals and interests of the perceiver” (Hoffman and Subramaniam 1995, 787); he summarized his finding in the claim that “people who think differently also, to some extent, see differently,” thus connecting the direction of viewers’ gaze to the purpose of a task given to them (Yarbus 1967, 211), and it is in this context that Brown calls Yarbus “the ‘godfather’ of eye-tracking studies” (2015). But several decades before Yarbus, Guy Thomas Buswell (1935) studied the perception of images showing different types of artwork and found that the content of pictures would produce regular patterns of eye movements among participants, who favoured certain areas over others, thus offering early evidence on the cognitive and perceptual operations taking place during a viewing task (Henderson, Weeks Jr., and Hollingworth 1999, 210). In a more recent article, Hoffman and Subramaniam refer to Keith Rayner and Alexander Pollatsek’s (1989) findings that the choice of a viewer’s next fixation is apparently guided rather than random (Hoffman and Subramaniam 1995, 787). Finally, Henderson discusses “gaze control,” i.e. the real-time, active directing of the gaze to fixate on informative areas when viewing a scene, “in the service of ongoing perceptual, cognitive and behavioural activity” (2003, 498). For Henderson, considering gaze control is important for three reasons: first, the active direction of the gaze towards useful information is connected to the visual task at hand; second, eye movements betray the way attention is allocated in a viewed scene, thus revealing the ways in which the inner attentional system works; and third, eye movements offer a direct source of behavioural patterns that indicate continuous visual and

cognitive processing (2003, 498). The present references are only indicative of the list of scholars that have addressed this issue.

This idea of selectivity during saccades and fixations has been under the spotlight in relevant research, regarding mainly the cognitive processes that potentially guide and control the gaze to form such patterns of selectivity. Given the different functions of the fovea vs. the peripheral retina, i.e. the ability for visual acuity vs. sensitivity to light and motion as described earlier, it is obvious that the two coordinate in order to guide the attention of the viewer so that the viewing task is, in a sense, optimized. According to Cohen, it is considered that the peripheral retina picks up potentially informative areas, for the fovea – with its economically selective sampling – to be subsequently directed to those areas so that they are seen in detail. The fact that these successive fixations create patterns stands as an “observable indication of selective processing, especially in visual scanning tasks;” and yet, Cohen notes, although widely acknowledged in the relevant literature, the mechanism mobilizing this process of selection is only partially comprehended (2017, 271–272). An essential element in understanding the complexity of this selectivity is the relation between “overt” and “covert” visual attention. The former refers to the movement that eyes perform when exploring a visual array, in order to bring to the centre of the fovea an interesting part of the scene so that it is clearly discerned; the latter, on the other hand, refers to a much faster, but hidden visual attention mechanism that processes interesting areas internally and is thus related to the direction that overt attention takes; consequently, covert and overt attention should not be studied independently of one another (Hoffman 1998, 119; Henderson 2003, 498). A saccade towards the next potentially interesting point of fixation may occur involuntarily due to an “exogenous” change in the visual features of the observed object, e.g. sudden movement, brightness or appearance of a stimulus; or voluntarily due to an “endogenous” cognitive operation connected to the task-related goals or expectations of the viewer which dictates, in a sense, the shift of attention (Hoffman 1998, 119–120; Hyönä 2010, 173; Remington, Johnston, and Yantis 1992, 279).³ Finally, there is “partial interdependence” between eye movements and attention, in the sense that attention is able to change direction regardless of eye movements, but the eyes require the change of attention before they move (Hoffman 1998, 119–120).

3 Exogenous and endogenous control of attention produce different types of saccades as well. Saccades that are the result of exogenous control (image salience) are called “reflexive saccades,” whereas the ones caused by endogenous control (the viewer’s internal decision-making mechanisms) are called “volitional saccades” (Dyer and Pink 2015).

The parts of a visual stimulus on which subjects fixate, therefore, appear to be determined by two factors, one that is generated by the visual environment itself and another one that is triggered by the perceiver's cognitive operations, which are considered to be task-related. The first one is based on "bottom-up" features, i.e. consisting of external characteristics visible in the scene itself; and the second one is a "top-down" factor, which refers to structures of knowledge that guide the direction of the eyes (Henderson 2007, 219; 2011, 596). During movie-watching, examples of bottom-up factors are characteristics of a scene that draw the viewer's attention, such as colour, lighting, edges, and motion; examples of top-down factors are the tasks that viewers have to perform, along with their individual preferences as well as mental formulations of a scene at the time of viewing (Hutson et al. 2017, 3).⁴ Dyer and Pink comprehensively summarize all these points by describing eye movements as "'bottom-up' processing when the eye makes reflexive saccades to salient stimuli within a scene, or 'top-down' when a viewer uses their volitional control to direct where the eye should look;" the authors also stress the importance of studying both these types of saccades "for understanding how we interacted with complex scenes in everyday life" (2015). Interestingly enough, Henderson also notes that research has only recently began to study ways of combining the two (2011, 596); the recent period that he refers to roughly coincides with a burst in commercial applications of eye-tracking research and the development of new tools (Płużyczka 2018).⁵ Film theory and studies on movie-watching in general can only benefit from such a coincidence, as it is a remarkable opportunity for them to be substantiated by actual gaze data that will incorporate eye movement and cognitive operations together. One of the main reasons for this claim, as stated at the beginning of this section as well, is the fact that research on various types of media has shown potential differences in the way information is obtained from them visually; the properties

4 Hutson et al. also note here that even within top-down processes there are those that are volitional, i.e. related to a goal that a viewer might have while watching a movie clip, and others that are more mandatory or unintentional, such as directing attention to the character who is speaking (2017, 3).

5 Płużyczka draws from iMotions, a company that provides eye-tracking research solutions ("Exponential Growth in Academic Eye Tracking Papers over the Last 40 Years!" 2011). Nevertheless, the exact numbers that she uses to describe the surge in academic publications on eye-tracking in the late 1990s and early 2000s might not be entirely accurate, since Google Scholar, in which the search was made, is not an exhaustive database of all research papers. In addition to that, iMotions states that only the term "eye tracking" was used for the search, but relevant research may be referred to by scholars using other terms such as "gaze data," "eye movement," etc. Nonetheless, the numbers remain indicative of the sharp rise in relevant work since 2000 or so.

of narrative film seem to produce different eye movements as well, which may be an indication of differences in the way those movements connect to cognition compared to e.g. reading (Hutson et al. 2017, 4). The following section reviews work that has been conducted on this consideration.

Eye-tracking and Filmmaking

Much of the research cited here explains how eyes work in general, a fact which imposes on movie-watching the same insurmountable constraints as it does to all visual tasks, and this is taken into consideration in eye-tracking studies that focus specifically on cinema. Brown discusses the fact that human eyes selectively pick up only a small set of available visual information through fixations, added to the inescapable “blindness” that occurs during saccades, sleep, and normal blinks; he stresses the necessity of this temporary blindness for survival, and that our vision is “deliberately, ‘perfectly’ imperfect” (2018, 20). Moreover, the phenomenon of saccadic suppression, i.e. the loss of sensitivity of human vision during saccades, rather than being a deficiency, in fact allows us to see clearly during fixations, as the world before our eyes does not blur in a distracting way while jumping from one fixation to another (Brown 2018, 18; Smith 2013, 168);⁶ this makes it “a structuring principle of vision itself,” in the sense that seeing and not seeing are not mutually exclusive but combined features of vision (Brown 2018, 18; Smith 2013, 168). Humans do not perceive this temporary blindness, as their visual experience of both the real world and cinematic action are similarly perceived as uninterrupted; and yet – due to saccadic suppression as well as the fact that a person’s field of vision can cover only approx. 3.8 percent of an average cinema screen – in reality viewers literally do not see as much as 96.86 per cent of a movie, which means that they only get to actually see roughly the remaining 3.14 per cent (Brown 2018, 21–22; 2015; Smith 2013, 168).⁷ The obvious consequence of these limitations is that they can potentially affect every aspect of film spectatorship out there; if the medium of film perception itself, the human eye, has fundamental functional constraints, then it is probably precarious to attempt any argumentation on film theory without adding those constraints to the equation.

6 In fact, saccadic suppression is easy to verify by our own inability to see our eyes moving in the mirror (Matin 1974, 899; Tatler and Trościanko 2002, 1403).

7 Brown further enhances his point (2018, 22) by also referring to Mary Ann Doane’s estimation that film viewers are sitting in the dark for approx. 40 per cent of a movie, due to the black leader intervening between film frames as they rapidly succeed one another in analog projections (2002, 172).

Apart from the way they affect the reception of movies, an interesting side-effect of these constraints with regard to eye-tracking research involves film creation as well, i.e. the creative decisions of filmmakers. Smith stresses how important it is for filmmakers to know exactly what viewers are looking at, especially given the extremely small percentage of a movie that viewers actually get to see (2013, 168). Naturally, the question arises whether filmmakers are in fact able to control what and where viewers see, and if they indeed do so. This question has been discussed over many years, both inside and outside the context of eye-tracking studies; but it is within the latter that a renewed understanding of human movie-watching behaviour has started illuminating not only how viewers see movies, but also how filmmakers have been creating them. Obviously, based on the discussion so far, the question can be rephrased as whether bottom-up features in the filmic image can be considered powerful enough to grab the viewers' attention in a way that overwhelms top-down processes; and at the same time, what role narrative comprehension plays in this.

Concerning the level of impact that bottom-up factors may have on viewing, eye-tracking analysis of movie-watching considers observations that have been made on vision, such as that of "attentional synchrony" in perception studies. The term refers to certain bottom-up features, e.g. motion being strong enough to collectively attract the attention of multiple viewers, with that attraction being observed at the same points in time during watching and on the same parts of the image, as demonstrated by the viewers' clustered fixations on these parts.⁸ Attentional synchrony is used for observing the intensity of the effect that exogenous features have on viewer attention mainly with regard to the preference that human vision may show for elements such as e.g. motion and human presence.

Nevertheless, attentional synchrony presents differences both between free and task-oriented viewing,⁹ as well as between different kinds of dynamic scenes, i.e. edited movies in contrast to natural, unedited clips.¹⁰ More specifically, it has been found that the existence of specific tasks or instructions given to participants

8 The element of temporal coincidence is important in understanding attentional synchrony. Although the bottom-up points of interest in an image that will attract viewer attention are generally predictable regardless of an image being static or dynamic, attentional synchrony is much higher in dynamic scenes. In static scenes viewers look at those points of interest at the same time far less frequently (Smith and Mital 2013, 1; Hutson et al. 2017, 2–3; Smith 2013, 170).

9 "Free" viewing in this context means that the viewer is not provided with specific task instructions.

10 Hutson et al. describe natural clips as those dynamic scenes that do not have "a narrative or any filmmaking techniques" (2017, 3).

during free viewing of dynamic natural scenes may override the influence of exogenous factors, thus decreasing the level of attentional synchrony (Smith 2013, 183–184; Smith and Mital 2013, 1; 20). The case with edited dynamic scenes, which concern the overwhelming majority of mainstream cinema production, is not that straightforward. There is some evidence that the existence of tasks given to viewers can indeed change the levels of observed attentional synchrony, but their effect is not constant throughout the duration of a movie; during the time viewers are watching a movie, the existence of a task competes with the effect of directorial decisions such as *mise-en-scène*, staging, and editing. It is obvious, therefore, that the effects of endogenous vs. exogenous factors while watching edited dynamic scenes – for instance, during narrative film viewing – has not been conclusively resolved yet, something that Smith himself also acknowledges (2013, 184–185).¹¹

Attentional synchrony is essentially a marker of the collected spatial and temporal interest of multiple viewers on parts of the image that are thus considered informative or semantically important, and eye-tracking can support theoretical assumptions with regard to that importance. According to Smith, saccade frequency as well as attentional synchrony have been found to peak immediately after a new shot is introduced, and decline afterwards, when the content of the shot becomes familiar to viewers (2013, 176). Acknowledging this fact and conducting further research today with contemporary eye-tracking technology can also support the foundations of an older theory by testing it with equipment unavailable at the time it was originally formulated. Julian Hochberg and Virginia Brooks worked in 1978 on the comprehension of non-overlapping shots in edited movies and used the term “visual momentum” to describe the phenomenon by which viewers rapidly sample the image on the display with selective fixations, constantly looking for landmarks that appear across successive views, which would help them put together a continuous mental image of the cinematic space shattered across edited shots. This “impetus to obtain sensory information, and to formulate and test a schema,” Hochberg and Brooks explain, helps viewers comprehend what they see on screen, and this is a “motivating factor” for them to maintain this foveal sampling in their continuous search for visual comprehension. Visual momentum actually explains why edited shots are considered more interesting to watch than continuous ones; in the latter, as soon as saccades scan the screen in rapid succession they exhaust the informative parts

11 Smith refers here to the work of Janna G. Spanne (2006) on eye-tracking and narrative fiction film.

of the image and visual momentum declines (Hochberg and Brooks 2007, 207–208). Hochberg and Brooks thus predicted when a scene becomes “cinematically dead,” i.e. presenting no further interest to viewers, which is also manifested in the frequency of the viewer’s glances (2007, 207–208), and they did so without having the eye-tracking technology that is commonly used today (2007, 209–210; 214–215).¹² As Smith notes, this means that film editors are in the position to control visual momentum by renewing the interest of viewers in a sequence of edited shots e.g. by reframing scenes or cutting to new ones when necessary (Smith 2013, 177).

The level of control exogenous factors have on viewing edited narrative movies in particular has attracted more attention in recent years. In fact, observing the level of attentional synchrony in movie clips, which is entirely feasible today with eye-tracking equipment, has been used as a measure of the extent to which the endogenous, cognitive-driven control of attention may succumb to the power of exogenous, visual control. Loschky et al. have made a series of important observations: they indeed re-affirm the fact that viewers attend differently to edited narrative scenes compared to both static scenes and dynamic unedited ones; in the latter two cases, endogenous control, i.e. providing viewers with a task, may very well override the exogenous ones, i.e. the visual attributes of a scene, in attracting viewer attention. But when viewers watch an edited narrative movie, the continuity editing style is powerful enough to produce and maintain high levels of attentional synchrony among viewers; and this is also true as filmmaking has evolved over the decades, using more intense editing, motion and visual contrast (Loschky et al. 2015, 1–3).¹³ In other words, despite the fact that viewers of edited movies are operating under the task of comprehending the narrative, which is an endogenous, higher-level cognitive operation, this task is not enough to produce differences in their eye movements on screen as it could do with static or unedited scenes. Loschky et al. use the term “tyranny of film” to refer to this overwhelming power that mainstream Hollywood moviemaking

12 As Hochberg and Brooks describe, in their experiments on visual momentum they used pictures in the form of slideshows instead of videos (2007, 209–210; 214–215); Smith, writing in 2013 and aware of the potential of eye-tracking, notes that Hochberg and Brooks lacked the kind of eye-tracking equipment available today (2013, 177), which is an indication of his trust in this technology.

13 Hutson et al. (2017, 21) also note in their article that their findings significantly agree with those by Loschky et al. Moreover, Loschky et al. cite David Bordwell’s work on the stylistic changes in continuity editing that have taken place over the past four decades or so, collectively called “intensified continuity:” Bordwell notices an increase in the rapidness of editing, the increasing use of lenses with more extreme lengths, the intensification of close-ups in dialogues, and the use of free ranging cameras (Bordwell 2002, 16–21).

techniques have to attract the gaze of viewers to specific parts of the scene, which probably allows very little room for differences in the way viewers behave visually when watching a movie (2015, 18).¹⁴

The impact of narrative comprehension seems to be an important issue in eye-tracking studies of edited films. Smith stresses that narrative is prominent among endogenous factors potentially affecting movie-watching, and possibly even inscribed in viewers' gaze patterns (2013, 185). In his own work on Attentional Theory of Cinematic Continuity (AToCC), Smith associates the term "continuity" as it has been used in film theory with "the continuity of viewer cognition: what they are attending to, what they are perceiving, and what they are expecting" (2012, 2). He describes cognitively active viewers, who are constantly attentive to a flow of usual elements – such as dialogue, off-screen audio, motion, gaze direction, gestures, etc. – as motivational cues for their attention to be guided naturally across the content of consecutive shots (2012, 2; 2013, 174). He used eye-tracking when formulating the AToCC, collecting gaze data from various types of shots, in order to demonstrate "how the continuity editing rules create the conditions necessary for continuity perception but also how films that do not adhere to the continuity editing rules can also result in continuity" (2012, 2). By doing so, Smith notes, the AToCC becomes a proposed way by which experimental research on vision can catch up with decades-long intuitions of filmmakers (2012, 23). Therefore, as exemplified by AToCC, eye-tracking is a key component in the methodology of this research and its importance goes both ways. On the one hand, common and established filmmaking methods such as continuity editing can become the inspiration for vision scientists to further explore perception by reverse-engineering those practices that have been successfully used in cinema for so long, but never received adequate scientific substantiation as to the physiological and cognitive grounds of that success. On the other hand, the same substantiation, once achieved, can provide some kind of formal codification for those filmmaking methods that can further benefit filmmakers: in the present example, beyond simply knowing that continuity editing enables viewers to follow action and narrative across cuts, becoming

14 In the experiment that Loschky et al. conducted, even when a test group of viewers with prior access to the context of the test clip was compared to another test group that was not given the same context, the authors indeed noticed an increased comprehension of the narrative in the former group, but the levels of attentional synchrony were only subtly different between the two groups. Surprisingly, as the authors note, similar results were observed in a subsequent experiment with a single long-take (unedited) clip, which they performed in order to test the general applicability of their initial findings (2015, 18–19).

aware of the perceptual basis as to why and how this works can make filmmakers appreciate more formally and systematically the tools at their disposal.

It is more than obvious that mainstream Hollywood cinema and the power of the continuity editing style are important enough to bring the concept of narrative comprehension into play, effectively forcing a separate line of eye-tracking studies just for their sake. It should be fairly obvious by now that, unlike the practice of simply observing and re-applying a successful mode of making movies, eye-tracking has begun shedding light on established observations on viewer attention by filmmakers and the drives behind that attention, in a way that it can better inform such creative choices. Eye-tracking provides filmmakers with “physical evidence” of the otherwise inaccessible information of viewers’ mental activity (Hutson et al. 2017, 2). This condition is not only fairly unprecedented in the history of moviemaking technique, but also probably the most groundbreaking one so far.

Research Directions

The extent to which eye-tracking studies can be widely applied to cinema is a matter that has met certain counterarguments as well. Brown comments in particular on the work of Smith in this field; he notes that Smith sidesteps the important issue of the universal applicability of eye-tracking studies, because of the latter mainly focusing on Hollywood mainstream cinema, thus also legitimizing one film form over others. In addition, Brown continues, Smith also downplays the practical effectiveness that movie-watching itself had both for testing of film theories and in the movie industry, long before the existence of eye-trackers. By doing so, eye-tracking is implicitly made to appear as the pinnacle of authority with regard to knowledge about film (2015). Nevertheless, there is risk in such positions that underestimate the power eye-tracking has in any study of perception, let alone film watching. It seems unreasonable to be offered with a tool that can help verify or falsify assumptions about visual perception of film in relation to cognitive operations and be suspicious of it on the basis that intuitions of skilled filmmakers eventually worked after years of trial and error. It is true that accurate findings require proper research methodology so that errors are eliminated and those findings should cover as much of the art of cinema as possible without being limited to Hollywood narrative cinema; but that is a condition that depends on the scope of researchers and the way that scope may change over time, and should not concern the validity of eye-tracking per se.

On the part of research itself, there is a difference between claiming universal applicability of findings and admitting that findings concern a specific category of cinema, in this case mainstream narrative cinema; the fact that a tool is used by researchers with specific scopes does not mean that the tool itself is to blame for the results of that research being applicable to a specific type of movies only.

Moreover, the popularity or dominance of a certain kind of film production may reasonably affect the scope of a significant part of academic research and the application of new tools, at least in its early steps. It is true that mainstream Hollywood-style films represent a considerable part of film production today; and with eye-tracking having started to be seriously used in cinema research in relatively recent years, it is natural to expect an initial tendency to focus on films with the most widespread impact. Nevertheless, eye-tracking research is already being carried out on a wider variety of film types. Tessa Dwyer and Claire Perkins, for instance, focus on what has been called “slow cinema,” i.e. non-mainstream films that have an internal “narrative, aesthetic and/or political preoccupation” with the concept of time. Films of this type often have unusual durations or a slow rhythm “characterized by static camerawork, minimal editing and scarce or slow movement within the frame” (Dwyer and Perkins 2018, 103). The authors discuss whether such films involve “a distinctive form of seeing – whether taking one’s time and experiencing the phenomenon of ‘dead time’ might facilitate an embodied, subjective mode of viewing” (Dwyer and Perkins 2018, 103) and acknowledge the contribution of eye-tracking in providing “a unique opportunity to examine these ideas at an empirical level” (Dwyer and Perkins 2018, 104). Among other things, Dwyer and Perkins consider long takes, a technique which is in itself worthy of more specialized research. The impact that long takes have on film viewing compared to edited scenes has attracted the attention of film theorists, but can be explored in different depth when gaze data from viewers is taken into consideration. The observation that long takes bring along a sense of continuous time and space in contrast to “the synthetic continuity of time that is achieved in continuity editing” (Gibbs and Pye 2017, 6) creates associations with the way reality is experienced, which clearly brings into play the way human vision works.

Eye-tracking has also been used in areas within film theory that one might consider to be less obvious, but are still significant in relation to the experience of movie-watching. One such area is the contribution of sound to the way viewers visually respond to movies. Dwyer and Perkins note that sound design is one of the additional factors, along with composition of shots and camera movement,

that need to be related to the length of shots so that visual momentum can be properly understood (2018, 120). Recent examples of relevant eye-tracking research include work on the way the direction of viewers' gaze can be affected by film music (Mera and Stumpf 2014), the combination of music, dialogue, and sound effects (Batten and Smith 2018), or even the availability and setup of sound equipment that is used during movie-watching (Korshunova et al. 2019). All this work is an important contribution to the wider literature of film theory, supported by an increased sense of credibility that the use of eye-tracking is capable of promising.

Closing Remarks

Although applications of eye-tracking in film theory have been a fairly new addition to academic research, results already appear to be encouraging. The insights that eye gaze data offer both to the filmmaking process itself and as support to formulations of theoretical questions are invaluable research assets today, especially after several years during which cognition-based film theories are gradually gaining more ground. Eye-tracking is a tool, which can facilitate the rapid acceleration of such approaches to film theory, and it is exactly the kind of acceleration that film theory needs, now that the cinema industry has entered a new era of production. Bearing in mind that new technologies of imaging and audience immersion – such as interactivity, 3D, augmented reality and VR environments, etc. – are gradually finding their way into filmmaking, the role of the viewer should probably be comprehended as thoroughly as possible in traditional moviemaking before moving into such media. Being able to decode and unlock the process of viewers' active engagement with the content of a movie can help film theory be more prepared when all these technologies will be fully incorporated in commercial cinema. With all its current shortcomings, eye-tracking can be of help in that direction.

The aim of this article has been to present an overview of the complexities of both the visual system and human perceptual mechanisms, and also to highlight the opportunities that technology provides to film theory today for tackling those complexities. The fact that cinema production is already showing signs that it outgrows practical intuitions and trial-and-error methodology is in itself a call for better substantiation and more accuracy of theory, which will hopefully communicate even more with filmmaking practice as well. With the ongoing improvement in eye-tracking technology, one can expect that even the last traces

of invasiveness during experiments will be eliminated, thus producing even more natural responses by viewers and more acceptable results. This is an opportunity for film theory to expand and evolve along with it.

References

- Batten, Jonathan P. and Tim J. Smith. 2018. Looking at Sound: Sound Design and the Audiovisual Influences on Gaze. In *Seeing into Screens: Eye Tracking and the Moving Image*, eds. Tessa Dwyer, Claire Perkins, Sean Redmond, and Jodi Sita, 85–102. New York: Bloomsbury Academic.
- Bojko, Aga. 2013. *Eye Tracking the User Experience: A Practical Guide to Research*. Brooklyn: Rosenfeld Media.
- Bordwell, David. 2002. Intensified Continuity: Visual Style in Contemporary American Film. *Film Quarterly* vol 55, no. 3: 16–28.
- Brown, William. 2015. Politicizing Eye-Tracking Studies of Film. *Refractory: A Journal of Entertainment Media* vol. 25 (February). <https://refractory-journal.com/brown/>. Last accessed 20. 10. 2021.
- Brown, William. 2018. In Order to See You Must Look Away: Thinking About the Eye. In *Seeing into Screens: Eye Tracking and the Moving Image*, eds. Tessa Dwyer, Claire Perkins, Sean Redmond, and Jodi Sita, 15–27. New York: Bloomsbury Academic.
- Buswell, Guy Thomas. 1935. *How People Look at Pictures: A Study of the Psychology of Perception in Art*. Chicago: The University of Chicago Press.
- Cohen, Karen M. 2017 [1981]. The Development of Strategies of Visual Search. In *Eye Movements: Cognition and Visual Perception*, eds. Dennis F. Fisher, Richard A. Monty, and John W. Senders, 271–288. London: Routledge.
- Doane, Mary Ann. 2002. *The Emergence of Cinematic Time: Modernity, Contingency, the Archive*. Cambridge: Harvard University Press.
- Duchowski, Andrew. 2007 [2003]. *Eye Tracking Methodology: Theory and Practice*. London: Springer.
- Dyer, Adrian G. and Sarah Pink. 2015. Movement, Attention and Movies: The Possibilities and Limitations of Eye Tracking? *Refractory: A Journal of Entertainment Media* vol. 25 (February). <https://refractory-journal.com/dyer-pink/>. Last accessed 20. 10. 2021.
- Dwyer, Tessa and Claire Perkins. 2018. Passing Time: Eye Tracking Slow Cinema. In *Seeing into Screens: Eye Tracking and the Moving Image*, eds. Tessa Dwyer,

- Claire Perkins, Sean Redmond, and Jodi Sita, 103–125. New York: Bloomsbury Academic.
- Exponential Growth in Academic Eye Tracking Papers over the Last 40 Years! 2011. *Imotions*. <https://imotions.com/blog/exponential-growth-in-academic-eye-tracking-papers-over-the-last-40-years/>. Last accessed 24. 04. 2020.
- Findlay, John M. 2004. Eye Scanning and Visual Search. In *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, eds. John M. Henderson and Fernanda Ferreira, 135–159. New York: Psychology Press.
- Gibbs, John and Douglas Pye. 2017. Introduction 1: The Long Take – Critical Approaches. In *The Long Take: Critical Approaches*, eds. John Gibbs and Douglas Pye, 1–26. London: Palgrave Macmillan.
- Henderson, John M. 2003. Human Gaze Control during Real-World Scene Perception. *Trends in Cognitive Sciences* vol. 7, no. 11: 498–504.
- Henderson, John M. 2007. Regarding Scenes. *Current Directions in Psychological Science* vol. 16, no. 4: 219–222.
- Henderson, John M. 2011. Eye Movements and Scene Perception. In *The Oxford Handbook of Eye Movements*, eds. Simon P. Liversedge, Iain Gilchrist, and Stefan Everling, 593–606. Oxford: Oxford University Press.
- Henderson, John M., Phillip A. Weeks Jr., and Andrew Hollingworth. 1999. The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing. *Journal of Experimental Psychology: Human Perception and Performance* vol. 25, no. 1: 210–228.
- Hochberg, Julian and Virginia Brooks. 2007 [1978]. Film Cutting and Visual Momentum. In *In the Mind's Eye: Julian Hochberg on the Perception of Pictures, Films, and the World*, eds. Mary A. Peterson, Barbara Gillam, and H. A. Sedgwick, 206–228. Oxford: Oxford University Press.
- Hoffman, James E. 1998. Visual Attention and Eye Movements. In *Attention*, ed. Harold E. Pashler, 119–153. Studies in Cognition. Hove: Psychology Press.
- Hoffman, James E. and Baskaran Subramaniam. 1995. The Role of Visual Attention in Saccadic Eye Movements. *Perception & Psychophysics* vol. 57, no. 6: 787–795.
- Hutson, John P., Tim J. Smith, Joseph P. Magliano, and Lester C. Loschky. 2017. What Is the Role of the Film Viewer? The Effects of Narrative Comprehension and Viewing Task on Gaze Control in Film. *Cognitive Research: Principles and Implications* vol. 2, no. 46: 1–30.
- Hyönä, Jukka. 2010. The Use of Eye Movements in the Study of Multimedia Learning. *Learning and Instruction* vol. 20, no. 2: 172–176.

- Jacob, Robert J. K. and Keith S. Karn. 2003. Commentary on Section 4 – Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, eds. Jukka Hyönä, Ralph Radach, and Heiner Deubel, 573–605. Amsterdam: Elsevier.
- Korshunova, Victoria, Gerard B. Remijn, Synes Elischka, and Catarina Mendonça. 2019. The Impact of Sound Systems on the Perception of Cinematic Content in Immersive Audiovisual Productions. In *2019 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR)*, 1–5. Ikoma: IEEE. <https://ieeexplore.ieee.org/document/8709163>. Last accessed 04. 06. 2020.
- Loschky, Lester C., Adam M. Larson, Joseph P. Magliano, and Tim J. Smith. 2015. What Would *Jaws* Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. *PLoS ONE* vol. 10, no. 11: 1–23.
- Majaranta, Päivi and Andreas Bulling. 2014. Eye Tracking and Eye-Based Human-Computer Interaction. In *Advances in Physiological Computing*, eds. Stephen H. Fairclough and Kiel Gilleade, 39–65. London: Springer-Verlag.
- Matin, Ethel. 1974. Saccadic Suppression: A Review and an Analysis. *Psychological Bulletin* vol. 81, no. 12: 899–917.
- Mera, Miguel and Simone Stumpf. 2014. Eye-Tracking Film Music. *Music and the Moving Image* vol. 7, no. 3: 3–23.
- Pluzyczka, Monika. 2018. The First Hundred Years: A History of Eye Tracking as a Research Method. *Applied Linguistics Papers* vol. 25, no. 4: 101–116.
- Rayner, Keith. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* vol. 124, no. 3: 372–422.
- Rayner, Keith and Alexander Pollatsek. 1989. *The Psychology of Reading*. Englewood Cliffs: Prentice Hall.
- Remington, Roger W., James C. Johnston, and Steven Yantis. 1992. Involuntary Attentional Capture by Abrupt Onsets. *Perception & Psychophysics* vol. 51, no. 3: 279–290.
- Smith, Tim J. 2012. The Attentional Theory of Cinematic Continuity. *Projections* vol. 6, no. 1: 1–27.
- Smith, Tim J. 2013. Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory. In *Psychocinematics: Exploring Cognition at the Movies*, ed. Arthur P. Shimamura, 165–191. Oxford: Oxford University Press.

- Smith, Tim J. and Parag K. Mital. 2013. Attentional Synchrony and the Influence of Viewing Task on Gaze Behavior in Static and Dynamic Scenes. *Journal of Vision* vol. 13, no. 8: 1–24.
- Spanne, Janna. 2006. Task Impact on Cognitive Processing of Narrative Fiction Film. Master's Thesis, Lund: Lund University. <http://lup.lub.lu.se/student-papers/record/1327791>. Last accessed 24. 04. 2020.
- Tatler, Benjamin W. and Tom Trościanko. 2002. Last but Not Least. *Perception* vol. 31, no. 11: 1403–1406.
- Yarbus, Alfred L. 1967. *Eye Movements and Vision*. New York: Springer Science.