Do You See What I See?

The influence of gaze cues on visual attention

Jules Eekelaar

ANR: 480176

SNR: 2017387

Master's Thesis

Communication and Information Sciences

Specialization: New Media Design

Faculty of Humanities

Tilburg University, Tilburg

Supervisor: Dr. R. Cozijn

Second reader: Dr. R.M.F. Koolen

June 2019

Acknowledgements

I would like to express my very great appreciation to Reinier Cozijn for providing me with the opportunity to write my master's thesis on this topic. From day one I've been thrilled to conduct research in this field and the collaboration with Cozijn made this even better. Additionally, I would like to thank both Reinier Cozijn and Ruud Koolen for their supervision and support during the realization of this work.

Abstract

Individuals have no trouble interpreting a sequence of different shots as a continues storyline when viewing a movie. To guide their attention and help them perceive continuity, directors use attentional cues. Different fields of research (i.e., static advertisement) explored how these cues influence the attention of individuals. However, relatively limited experimental studies on the use of attentional cues in movies have been conducted. The current study investigates how movies are able to direct visual attention by the use of the gaze cue. Guided by the *Attentional Theory of Cinematic Continuity* of Smith (2012), an eye-tracking experiment (*N*=49) with a 2x2 (Congruency vs. Display time) within-design was conducted measuring recognition scores and viewing times. The present study showed that for a Display time of 800ms, the gaze cue positively influenced the recognition scores of participants. This supported H1. Additionally, the gaze cue did not significantly influence the participant's viewing time on objects. However, again for a Display time of 800ms, results were in line with findings of the recognition task and H2. Lastly, objects in the congruent condition were further away from the person initiating the gaze. This might have influenced the results and could be a topic worth investigating in future research as well as the effect of a gaze with head rotation.

Keywords: gaze cue, Attentional Theory of Cinematic Continuity, visual attention, eye-tracking, movies

Table of contents

Acknowledgements	2
Abstract	3
Introduction	**6**
Literature review	**8**
2.1 Defining continuity	8
2.2 Perception	10
2.3 Attentional Theory of Cinematic Continuity	11
2.4 Perception and visual attention	13
2.5 Attentional orientation	14
2.6 Attentional Cues	15
2.6.2 Gaze cues	16
2.7 Eye-tracking	18
2.8 Previous studies on gaze-cues	19
2.9 Research question and hypotheses	21
Method	**23**
3.1 Participants	23
3.2 Apparatus	24
3.3 Materials	24
3.4 Design	27
3.5 Instrumentation	27
3.6 Procedure	28
3.8 Data Analysis	30
Results	**33**

4.1 Recognition analysis 33

4.2 Eye-movement analysis 35

Discussion 37

Implications 42

Limitations and future research 42

Limitations 43

Future research 43

References 45

Appendices 49

Appendix A 49

Appendix B 51

Appendix C 52

Appendix D 53

Appendix E 54

Appendix F 55

Appendix G 56

Do You See What I See?

Dating far back in time, the religion of Egyptians was aimed at survival after death. In their eyes, one was able to live forever if their physical body was preserved after their death. This ultimate goal of creating immortality by freezing time existed throughout the ancient centuries. Much later, King Louis XIV did not choose to be mummified. He decided to have himself eternalized in the form of a painting. Today, images support remembering events. Bazin (1967) describes how images are used to replicate or change a reality that can be preserved forever. Whereas paintings always include the slightest of subjectivity and skill of the painter, the absence of human involvement in the process of photography obsessed many. The capturing of pictures is entirely objective. Film theorists like Bazin (1967) and Kracauer (1997) stated that the motion picture (also known as, film or movie) was able to transcend the power of photography by being able to capture a sequence of time instead of a fixational point. The creation of motion pictures was mainly due to an important collection of photographic elements that were later combined with novel, film-specific characteristics (i.e., editing and sound). Additionally, both authors described how the motion picture allowed to be immersed in the captured reality that it established. Similarly, Anderson (1998) suggested that viewers of motion pictures experience the picture as if they were there themselves. Moreover, the perceptual processing of these motion pictures is said to be the same as processing real life. Contrastingly, even though this illusion is supposed to be perceptually identical, viewers do not feel the tendency to run when shown something scary (Carroll & Seeley, 2013).

A motion picture tells the viewer a story using different cinematic sequences, consisting of various shots. Shots are made by a collection of individual frames. Starting with the smallest part, frames are photographic stills that, when shown in a fast sequence, create the illusion of motion. Work by Sekuler (1996) explains how the original German gestalt psychologist Max Wertheimer (1912) calls this phenomenon apparent motion. Moreover, the eyes are unable to see the difference between these forms of apparent motion and real motion (Hildreth & Koch, 1987). A shot is an uninterrupted camera take consisting of a predetermined number of frames per second (i.e., 24 fps), running for an undetermined period. Multiple shots together form a cinematic sequence. These sequences are used to tell a story, highlight specific actions, events, and point of views (Carroll & Seeley, 2013).

The process of connecting consecutive shots is called editing, with the single connection between two shots named an edit. There are multiple different edits, but the most basic one is a cut, also known as the director's cut (Shimamura, 2013). These techniques enable the creation of a coherent motion picture and a continuous storyline. Surprisingly, viewers can easily understand how a sequence of shots belongs together and forms a story. One might think that this is all because of the way we learned to interpret movies. However, as will be explained in the following sections, this is due to the fascinating interaction between how perception/attention works, and how filmmakers learned to make use of perceptional and attentional processes in order to guide them, or even trick them. More specifically, filmmakers have learned how to guide attention and trick perception by the use of attentional cues. The film domain is not the only domain using these cues to guide attention. These cues have also been explored in advertising and marketing domains. However, experimental research in the film domain still offers great potential.

This thesis will, therefore, go in depth on how movies direct attention by the use of these attentional cues. Additionally, an experiment will be conducted to test the effect of these cues. Specifically, the gaze cue. All in order to provide an answer for the main question: How do movies direct visual attention? To provide a solid theoretical background, the following sections will focus on how the need for a continuous editing style evolved, how, by the use of these cues, attention can be guided in order to perceive a continuous edit, how the gaze-cue, and different cues, work, and what has been performed on this topic.

Literature review

2.1 Defining continuity editing

An important point in time was the discovery of new ways of manipulating motion pictures. At the end of the 19th century, during the "primitive" period, George Melies was one of the first, by accident, to systematically manipulate the motion picture, and thereby shaping many techniques that are still used today. Being a stage magician by origin, he introduced techniques like in-camera disappearing by discovering how to stop a shot mid-take and continue after having changed the stage. His work sparked the popularity and development of film editing. In the early 20th century, the first films were produced that consisted of a narrative using multiple sequential shots (Fischer, 1999). Additionally, as innovative techniques such as the close-up and cross-cutting were discovered, the primitive period came to an end (Bordwell, Staiger, & Thompson, 1985). To elaborate, close-ups are very narrowly framed shots of the subject that are used to show more detail. The cross-cutting technique is often used two different events are shown in parallel to indicate that they took place in the same time frame. Editing became the most important aspect of film making (Kracauer, 1997; Pudovkin & Montagu, 1958).

In the period that followed, movie techniques changed completely. Techniques that were barely used in the primitive period, such as the cut-in, point-of-view, or eyeline matches, were now extensively used to construct time and space (Bordwell et al., 1985). Cut-in shots are close-ups of something visible in the main shot. A point-of-view shot shows what the subject is looking at. These shots are often preceded by a shot of the subject looking a specific direction. Similar to the point-of-view shots, eyeline matches are shots that show where the subject is looking. However, the preceding shot shows that the subject suddenly looks in a certain direction. This is the cue for the second shot to present the object of attention.

Single shot films evolved into multi-shot narratives. Movies that had great success during that time had one corresponding characteristic: all of them included “continuity”. Continuity is the technique to merge the independent collection of shots into a whole that tells the audience a coherent story which is interpretable in a single way. This definition forms one of the basic principles of the Hollywood film (Bordwell et al., 1985).

In the 20th century, two highly contrasting styles emerged. Beginning with the “Classical” Hollywood style that evolved during the beginning of the 20th century and followed by the “New” Hollywood style that developed around the 1960s. Classical Hollywood is built upon continuity editing. Continuity within this style focused on creating both a coherent sequence of continuous shots and a continuous narrative. To establish this, changes in camera position and movement provided the audience with multiple perspectives on the story to be told and the highlights to be noticed. A typical scene within this classic style can be recognized as a strictly organized setting in which all actions have a clear drive and reason (King, 2002). Later, a “new” trend was set, breaking with the classical rules of filmmaking. It was named “New” or “Renaissance” Hollywood for a number of reasons, such as influences of American events during that time (i.e., “black power,” the Kennedy assassination), and the rise of large film-production companies. A scene in this era can be recognized as more stylistic. Classical rules such as “the clear goal of the main character” were allowed to be broken in order to create novel experiences (King, 2002). To elaborate, in classical movies, every event takes place because of clear reasons that precede the event. That makes the movie narrative very easy to understand (Carroll, 1985). Despite the lack of these classical elements within this new style, the element of continuity between shots remained rather untouched.

As mentioned above, according to Bordwell et al. (1985) continuity is the result of a seamless collection of independent shots that together form a coherent whole. In order to guide an individual's narrow focus, techniques like close-ups are used to highlight certain essential details within frames by using the movement of the camera (pan, tilt, track, etc.), zooming. This is called variable framing (Carroll & Seeley, 2013). However, making one continuous shot has proven to be impractical. This introduces the need for editing techniques. To edit in a way that is easy to follow for the viewer, one has to follow the rules of continuity editing to make these sequences of shots naturalistic to perceive (Smith, 2012).

2.2 Perception

Berliner and Cohen (2011) described how each of the most critical stages of real-world perception shows similarities with the perception of film. Contrary to common belief, real-world perception is constructed by combining non-continuous images. During the sensory input and encoding phase, a coherent representation of the surroundings is created by cognitively combining small parts of visual information that are collected using independent fixations. Additionally, humans make predictions based on common patterns and continuously try to identify causal relationships. Hence, the way spatial representations of film are created is identical to the way the real world is perceived, as the same principles apply when during the encoding of the perceived images.

The process of perception does come at a cost since the brain is very aware of inconsistencies in continuity. Berliner and Cohen (2011) explain this using the term "accommodation ranges." It describes how humans have a certain level of acceptance that monitors sensory input. For example, the accommodation range for perceiving motion in film is 24 frames per second. If a clip is played below that rate, the lack of fluid motion is noticed immediately. Therefore, filmmakers should attend to provide shots, and the sequence of shots, within accommodation ranges of the viewers perception. This will result in the viewer perceiving the highest continuity.

2.3 Attentional Theory of Cinematic Continuity

Smith (2012) gives his definition of accommodation ranges within the spectrum of continuity. With his *Attentional Theory of Cinematic Continuity* (AToCC), he defines how events in the world are assumed to be continuous, without claiming that perception is continuous. He proposes a theory that focuses on establishing continuity by guiding the viewers' attention rather than creating a full spatiotemporal representation of the scene. He calls this "a priori continuity".

While looking at a shot, the viewer creates expectations based on features within this shot. To make the viewer preserve (a priori) continuity, one should steer their attention to important features within a shot that helps them prepare for the next shot. If this is done successfully, the viewer's expectations will be matched in the next shot, a priori continuity will be perceived, and the cut will not be noticed. To illustrate, the act of an actor to suddenly look in a particular direction raises the expectation in the viewer that there is something worth looking at. If the following shot after the cut matches that expectation, attention will shift to the object at hand in a smooth manner, and a priori continuity is perceived. Contrastingly, mismatches also happen. In that case, the following shot does not match the viewer's expectations, or the viewer is not prepared for the cut. The change between the pre- and a post-cut clip will be noticed and interpreted as non-continuous. This puts the viewer in a confused state as he has to find pieces of information in the new shot that explain the relation to the previous shot in order to reconstruct his perception of continuity. This phenomenon is called a posteriori continuity. Additionally, this steers attention to the cut and pulls the viewer out of his immersive flow.

In order to minimize a posteriori continuity and maximize a priori continuity, viewers' attention should be guided in the most effective fashion. To do so, attentional cues are used. These are attention magnets that, either consciously or unconsciously, steer the viewers' attention to the preferred part of the scene. If a cut is implemented within this shift of attention, it has the least chance of being perceived by the viewer. This helps the viewer experience a continuous event rather than an incoherent sequence of shots. There are many different attentional cues like off-screen sounds, pointing gestures, attracting attention with motion, gaze cues, and so forth. All these attentional cues evoke different expectations on what will be seen after the shift of attention.

The AToCC consists of three main stages to guide the viewer's attention. The first stage focuses on attending to the shot. It describes how representations are created by locating a handful of objects of interest. Working memory is limited to holding a representation of more than four objects at a time. Additionally, these representations disappear in a very fast manner if they are not in the area of focus (Kahneman, Treisman, & Gibbs, 1992). Therefore, it may be assumed that knowing where viewers focus their attention on links to what objects are currently stored in their working memory. This is useful to know what object viewers use in a scene to track continuity. The second and third stages are cueing a cut and matching expectations. These stages focus on how different attentional cues prepare the viewer for a shift in attentional focus and provide a matching expectation to ensure a priori continuity (Smith, 2012).

2.4 Perception and visual attention

The way humans move their eyes to orient themselves is based on fixations and saccades. Fixation can be defined by the static focus of the eyes on an object of interest. To move a fixation to new stimuli, the eyes move. Such a move is called a saccade. During these saccades, it is impossible to perceive. This is known as saccadic suppression (Matin, 1974). The constant need to make saccades derives from the inability to perceive a large clear field of view at once. To elaborate, while the retina is able to visually absorb large parts of a scene at once, only a tiny portion of this field of view is highly detailed, sharp, and clear.

The retina works as a visual tube ranging from highly detailed in the middle to very blurred on the sides in three regions. The first region is the central vision, called the fovea. This region is very sharp but also very small, using only 2 degrees (from the center) of the total field of view. Moving further to the side on the retina, the parafoveal region can be found. This region is significantly less detailed than the fovea region but ranges to about 5 degrees from the foveal region. The last region is the peripheral region. This region is even worse in providing a detailed representation of reality and occupies the remaining field of view. As a result, the eyes continually align objects within the fovea's field of view to perceive them in detail (Rayner, 2009).

Because peripheral vision is so inaccurate, filmmakers make use of it to hide cuts in order to perceive continuity. In an experiment done by Simons and Levin (1997), participants were asked to watch a short film in which nine continuity errors were implemented without telling the participants or asking them to search for them. In the scene consisting of two women having a conversation at a dinner table, objects in the scene such as their dinner place or clothing objects were replaced or changed in color during the scene. In the initial viewing of the short film, none of the participants noticed any of the changes made in the short film.

Even in the second viewing of the film, only two out of nine continuity errors were noticed on average.

This phenomenon of being unable to detect changes in a scene is called change blindness and is studied by many researchers including Simons and Ambinder (2005). In their study, they describe multiple explanations for the occurrence of the phenomenon. These explanations mainly orbit around the idea of whether or not the object of change received enough attention to be noticed. For example, if the center of attention was on two women having a conversation, objects that were far away from the center of attention were noticed the least. This can be explained by objects in the peripheral vision being too vague for small changes to be noticed. Attentional cues like the match-action make good use of this phenomenon in combination with saccadic suppression to hide a cut. However, peripheral vision can also be used to guide the saccadic movement prior to fixation.

2.5 Attentional orientation

A well-known study that examined the relationship between vision and attention is that of Posner (1980). In his work, he proposes a simple model task to measure people's ability to orient and direct attention. To understand his concept of orienting attention, two important concepts should be defined. Namely, "covert" and "overt" attention. Covert attention is attending to something that is not currently visible. In other words, paying attention to something without moving your head and eyes in that particular direction. While your fixation stays in place, you pay attention to details in your peripheral field of view, for example, "looking" at your phone while walking but paying attention to obstacles. Overt attention is both paying attention and fixating on that particular object using the clear foveal region. Using the same example, this would mean looking at your phone while walking and looking up now and then to see if you don't bump into something.

The model Posner created was based on a simple stimulus detection task. Posner uses the term “detecting” to indicate a state of consciously perceiving the stimulus and being able to report its presence by pressing a button. Participants were asked to fixate on a small plus sign in the center of the screen. After fixation, an endogenous cue is provided in the form of an arrow for a short period of time, followed by a stimulus left or right of the center of focus. Endogenous cues are cues that are placed in the area of overt attention. Compared to exogenous cues, that are placed in the area of covert attention. The independent variable was the congruency or incongruency of the position of the stimulus and the direction of the arrow cue. The dependent variable was detection time, operationalized by measuring time to press a button. The study showed that, by using this model, Posner was able to effectively measure the movement of attention in the viewer’s visual field. Furthermore, it showed how viewers direct attention to the target stimuli prior to using their eyes and that this improves reaction time in the congruent condition (Posner, 1980).

2.6 Attentional Cues

The arrows used in Posner’s experiment also belong to the cues that direct attention, as explained in this section. In general, cues can be divided into two categories of attentional guidance. Namely, cues that attract attention and cues that direct attention. Cues that attract attention use the visual blur during eye movements to hide a cut. This optical blur can be seen as some an open-eyed blink as humans are not able to perceive during saccades (Matin, 1974). Cues that attract attention make use of the inability to perceive during these saccades to minimize the effect of a cut. An attentional cue that is closely related to this phenomenon is the match-action cut. Cues that direct attention prepare the viewer for a change in the shot. This technique provides the viewer with deictic cues to create a perceptual question in their mind. As explained by Gregory (1961), deictic cues are pieces of visual information that raise the perceptual question in the viewers' mind.

Such a cue can be an individual's gaze. Shepherd (2010) refers to this as "the deictic gaze." In other words, it can be seen as "the gaze that contains more information than just a direction of sight." He explains that the primary goal when following a deictic gaze is to establish shared attention or mutual mental state. Hence, he states that following an individual's gaze seems to be of fundamental importance for the development of advanced social skills, such as the Theory of Mind. Additionally, the development of this gaze-following behavior starts at an early age. Infants at two months old show signs of gaze-following behavior, with a high-frequency increase after seven to 10 months.

In the movie domain, this deictic gaze is often used to prepare viewers for the next shot (to preserve a priori continuity). For example, if an actor suddenly gazes in a particular direction, his gaze is a deictic cue that proposes the perceptual question "what is he looking at?". Viewers will try to solve this by following his line of sight. This enables the editor to introduce a cut and show the answer in the next shot. When directing attention, it is vital to shape the right kind of perceptual question in the mind of the viewer in order to minimize the noticeability of the cut. (Smith, 2006). The current study will focus on the gaze cue in particular. Furthermore, this cue will be used in the experiment to explore the guidance of visual attention on participants.

2.6.2 Gaze cues. As introduced in the previous sections, the gaze of the subject serves as both a deictic and endogenous cue that provides the viewer with the preferred kind of perceptional question. Gaze cues exist in different forms, either a gaze including the rotation of the subject's head or a gaze using only eyes excluding head rotation. Of both will be given an example. An example without head rotation can be seen in the movie "A Few Good Man" by Rob Reiner. Figure 1 shows an explicit gaze cue in this movie.

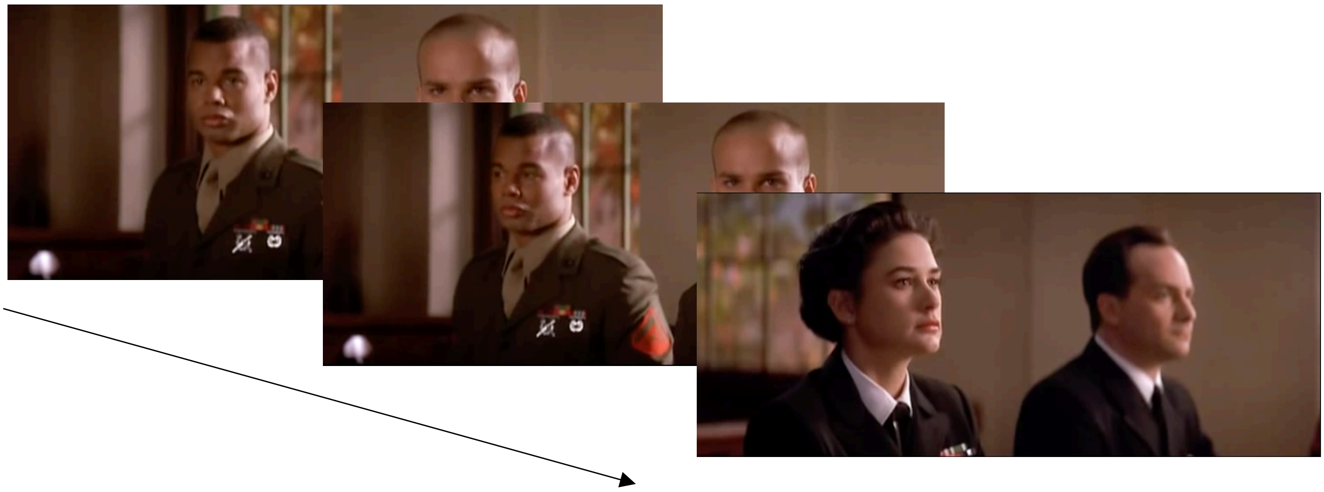

Figure 1. Frame 1 (left) shows the initial shot with two individuals looking in the direction of the camera. Frame 2 (middle) shows the left individual gazing to the left, introducing the deictic cue and raising the perceptual question: What is he looking at? This will induce an attentional shift following his line of sight. Frame 3 (right) answers this question and supports a priori continuity by matching the viewer's expectations. Noteworthy is that in this case, the gaze is done using only the eyes (A Few Good Men, 00:22:29).

An example of a gaze cue, including head rotation, can be found in the popular TV show "Suits." Figure 2 shows how the gaze cue occurs.

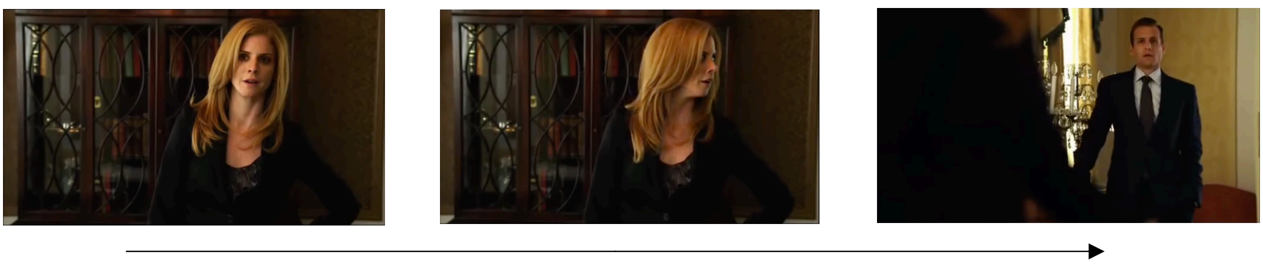

Figure 2. Frame 1 shows the initial shot of a woman looking into the camera. Frame 2 shows how she gazes to the right as she moves her head, again introducing the deictic cue and

raising the perceptual question: What is she looking at? This results in the viewer's attentional shift to the right. Frame 3 answers this question and supports a priori continuity by matching the viewer's expectations. Important is that in this case, the gaze includes a rotation of the head (Suits Season 1, Ep 1, 00:22:29).

2.7 Eye-tracking

So far, multiple theories on film have been discussed, visual attention and their relation. For example, it is assumed that attention shifts congruent to the gaze of the subject. It is mandatory to test these assumptions to make valid claims. In many fields of research, including film, a commonly used technology to do so is called eye-tracking. As the name suggests, it records the movement of a viewer's eyes while watching content relevant for the researcher. Over the years, researchers have moved from highly invasive techniques, such as attaching a small suction cup to the participant's eye, to non-invasive techniques. The new techniques use reflection on the eyeball as a reference for eye positioning. Distance is measured from the center of the pupil to the gleam of light reflecting on the eyeball. Luckily, current technologies involve the use of infra-red light that is invisible to the human eye compared. This technique is capable of producing highly accurate measurements.

Contrary to common belief, a film viewer is highly active during the event of watching a movie. As explained earlier, the inability to view a complete scene in one take results in many saccades per second to fully perceive the scene in detail. Multiple successive saccades and fixations together form a scan path. This is a time-stamped recording of the viewer's overt orientation during a period of viewing. Knowing that covert attention mainly shifts to locations preceding a saccade, eye-tracking data can provide information about the locations processed by the viewer with reasonable certainty. Within the short time, the viewer has to scan a shot, the scan path can provide the researcher with information about the choices the viewer makes to attend to the scene in the given time. If the scan paths of multiple viewers show high similarity, one can say that the shot, or sequence of shots, show attentional synchrony. Reasons for this phenomenon to occur is most likely due to shot composition and editing as findings showed that attentional synchrony was significantly more present in Hollywood movies compared to naturalistic registrations (Smith, 2013).

A relevant example of an eye-tracking study is the work by van de Schepop (2018). In his work, he investigated how a commonly used attentional cue, the match-action cue, is used to guide the viewer's attention. To do so, using eye-tracking and a recognition task, he measured the influence of the match action cut. Van de Schepop used two independent variables (Object location and viewing time) to measure recognition. Participants of his experiment were shown 20 movie clips in which five clips belonged to each of the four conditions in a within-subjects design. By showing objects in the congruent and incongruent field of the match-action cue, and showing these objects for a short (200 milliseconds) or long period (800 milliseconds) of time, he was able to show the effects of covert and overt attention induced by the match-action cut. The present study will use a similar experimental setup to test the research question and hypotheses that will be explained in section 2.9.

2.8 Previous studies on gaze-cues

The previous paragraphs showed how attention can be influenced by the use of editing techniques. More specifically, how gaze cues can direct attention in order to perceive continuity in film. However, not only film uses these techniques of directing attention by the use of gaze cues. Numerous studies have been conducted that focus on the influence of the gaze cue on human attention measured by eye-tracking. For example, in a study by Kuhn, Tatler, and Cole (2009), participants were asked to watch a magician doing a rather simple magic trick. They looked at a video recording of a magician who performs a trick in which he makes a lighter disappear. Their goal was to find the hidden event that would explain the trick. That is, dropping the lighter during the switch from one hand to the other. Two conditions were created, one condition in which the gaze of the magician pointed towards the hand incongruent to the area of interest (the space where the lighter is dropped), and one condition where the gaze of the magician is pointed in the direction of the hand that drops the lighter, congruent to the area of interest. Results showed that there is a significant influence of the gaze cue on the direction of attention. Participants that viewed the magician gazing towards the congruent area of interest were far more likely to observe the hidden event of the trick compared to the participants that were cued towards the incongruent area of interest. Most participants that detected the hidden event were already overtly fixating within this area. However, two participants that were fixating on the incongruent location did notice the lighter drop. This could indicate covert attention that played a role in the detection of the event. Something worth mentioning is that while the results are in line with most findings within this field, the experiment that was conducted included only 30 participants (15 per condition) so this study might not be up to standards to claim generalizable effects.

Compared to this naturalistic experiment of directing visual attention, other studies focused on the influence of gaze cues in the marketing domain. The study by Hutton and Nolte (2011) investigated whether the gaze of the model in static advertisement (either mutual gaze towards the viewer or an averted gaze towards the product) influenced the viewer’s attention. Results showed that viewers looked significantly longer at the product when the model’s gaze was also looking at the product. These types of studies have been conducted many times in the past decade. Another recent experiment that investigated cues on static advertisement, is the study by Pileliene and Grigaliunaite (2018) in which they tested the power of different visual cues, such as the gaze cue and the pointing gesture. Results showed that, as expected, advertisements without a cue score the lowest in all experiments. Gaze cues scored significantly better than the no-cue condition in all experiments, but not as good as the pointing gesture.

Similar effects can be found in the domain of online marketing on webpages. As these are also static advertisements, it might be expected to perform in a comparable manner. Similar to the study by Hutton and Nolte (2011), Sajjacholapunt and Ball (2014) conducted an experiment that investigated the different effects of the mutual and averted gaze on advertisements. Furthermore, they included changes (horizontal vs. vertical) in the orientation of the banners and measured memorability of the brand messages by the viewer. Results showed that, independent of the orientation of the banner, the conditions that included an averted gaze towards the brand showed an increase in attention and memorability by the viewer.

Not only humans but also numerous animals have shown to be heavily influenced by the power of the averted gaze. Itakura (2004) conducted a review on multiple studies that researched the effect of the gaze cue on different species. Not only the apes but also animals like horses and dogs showed significant performance in the experiments investigating gaze-following and joint attention. These studies show that gaze-cues are influential in many different domains.

2.9 Research question and hypotheses

Within the field of motion pictures, gaze cues are one of the numerous ways of attracting and directing attention in order to preserve cinematic continuity. Results of this experiment could strengthen findings by previous studies on gaze cues in the movie domain (i.e., van de Schepop, 2018), or in the advertising domain (Hutton & Nolte, 2011; Pileliene & Grigaliunaite, 2018; Sajjacholapunt & Ball, 2014). Additionally, considering the attentional theories of covert and overt attention by Posner (1980), Smith's (2012) AToCC on perceiving a priori continuity, this study could show results supporting both of these theories to strengthen their work further. Lastly, by contributing to the knowledge about how filmmakers use these attentional cues, both filmmakers themselves and others can learn how to better understand, apply, and explain the use of these attentional cues. The following research question is proposed:

RQ: *How do gaze cues direct visual attention when viewing movies?*

To operationalize the research question, a short introduction to the experiment should be given. When viewers watch a movie clip, including a gaze-cue, it is assumed that their attention is guided in the direction of the gaze. To test this in the current study, participants will be shown movie clips that include a gaze-cue in the end, followed by an object for the recognition task. Because the effect of the gaze cue might direct the attention of the viewer along the gaze, placing objects inside or outside this field proposes the assumption that objects inside the field will be recognized better. This results in the first hypothesis:

H1: O*bjects shown in the congruent location cued by the gaze cue will have higher recognition scores compared to objects in the incongruent location.*

To test if overt/covert attention is influenced by the gaze cue, different on-screen times for objects will be included. To elaborate, the gaze-cue might influence covert attention, overt attention, or both. In this experiment, a distinction is made in the on-screen time of objects to examine whether which one, or both, are influenced. By showing an object for 200 milliseconds (ms), it can only be seen if viewers are guided by covert attention. If the object is shown for 800ms, it can be seen by both the covert and overt attention of the viewer.

In the experiment of van de Schepop (2018), This was shown to be a valid way of measuring visual attention. Additionally, accompanied by the location of these objects (congruent vs. incongruent), this results in the following hypothesis:

H2: *Objects in the congruent location cued by the gaze cue will be viewed longer than objects cued in the incongruent location.*

Method

To investigate the effect of the gaze cue on visual attention, an eye-tracking experiment was conducted. This experiment, briefly mentioned above, consisted of a viewing task, a distraction task, and a recognition task. First, the participant was shown 20 short movie clips that each featured a gaze cue at the end. After each clip, an object was shown that differed in Display time (200ms vs. 800ms) and Congruency (congruent with gaze cue vs. incongruent with gaze cue). After completing the viewing task, the participant was asked to fill in a distraction task that focused on collecting demographical. Finally, a recognition task was presented. During this task, participants highlighted all the objects that they recognized from the 20 clips that they had viewed.

3.1 Participants

In total, 49 participants (18 men and 31 women, with a mean age of 21.6 years) participated in the experiment. Forty-one were sampled from Tilburg University's Human Subject Pool (HSP), and eight were students from Tilburg University sampled using convenience sampling. Participation in this experiment provided the participants from the HSP with 0.5 in HSP credit. To participate in the experiment through the HSP, individuals were required to have no difficulty reading English on-screen text. Additional participant data (i.e., whether or not participants wore glasses during the experiment) and demographical data were gathered in the distraction task. Ten people (21.7%) wore glasses during the experiment. The familiarity question in the distraction task aimed to detect whether participants had already seen the clips used in the experiment. In total, out of the 49 participants, the data of 39 participants were included in the analysis. Seven participants were excluded due to having bad data, and three participants did not pass the calibration test for the eye tracker.

3.2 Apparatus

Eye-tracking equipment located in the DCI Lab at Tilburg University was used for this study. Specifically, participants' eye movements were tracked with a SMI RED 250 eye tracker with a speed of 250HZ and an accuracy of 0.5º. A Dell 22-inch monitor with a resolution of 1650x1050 was used to provided image to the viewer combined with a Shuttle XPC Mini Desktop to run SMI Experiment Center to record the data. Participants were equipped with a Sennheiser Headset PC 320 to enhance immersion and provide sound when watching the movie clips while canceling out environmental noise.

3.3 Materials

The stimuli that were used in this study are 20 individual movie clips and 40 objects. The 20 movie clips were distilled from a collection of professional Hollywood movies and TV series such as Batman The Dark Knight, 8 Mile, The Accountant, and The Bank Job. The full list of movies used, including timecodes of the scenes, can be found in Appendix A. Each clip contained a gaze cue on the end of the clip, was 30 seconds long on average, and played at a framerate of 24 frames per second. After the clip, an object was added to one of the four conditions of Congruency and Display time. Concerning the lower accuracy of the eye tracker on the sides of the screen and the resolution of the Dell 22-inch Monitor, the clips were rendered on a 1650x1050 resolution in which the center 1280x720 pixels showed the movie clips. The object at the end of each clip was shown in either the congruent location or in the incongruent location for a short 200ms or a long 800ms. Figure 3 and Figure 4 will show an example of both the congruent and incongruent conditions.

Figure 3. Screenshots from the congruent condition. The left frame shows the movie clip that ends with a gaze cue towards the right. The right frame shows the appearance of the object (sushi) at that position, congruent to the gaze-cue's direction which is also to the right (The affair, 2014).

Figure 4. Screenshots from the incongruent condition. The left frame shows the movie clip that ends with a gaze cue towards the right. The right frame shows the appearance of the object (taco) at the position that is incongruent with the direction induced by the gaze-cue (The affair, 2014).

The selected gaze cues adhered to the following criteria. These consisted of an individual in the shot gazing towards a particular direction just before the cut. For types of gazing, both gazes with and without head rotation were accepted. However, gazes without head rotation were preferred. This was similar to the setting of the shot. Even though multiple people in the shot were accepted, shots with only one person were preferred. The shot that followed showed what the individual was looking at. Additionally, the on-screen position of the individual inducing the gaze cue and the area of interest in the following shot had to differ by a large enough amount to generate meaningful eye-tracking data. At last, the collection of clips that were selected needed to consist of gazes in all directions of the screen.

Second, the 40 objects were selected out of two icon packs downloaded from an online icon database (https://www.flaticon.com). The objects used in the experiment were scaled to a 220x220 resolution. To enhance the validity, the objects were decolorized to exclude the effect of color salience in icons that might influence the attraction of attention. The position of the objects was congruent with the direction of the gaze or incongruent with the direction of the gaze and differed across all clips so that each location on screen was covered. An example of some of these icons can be seen in Figure 5.

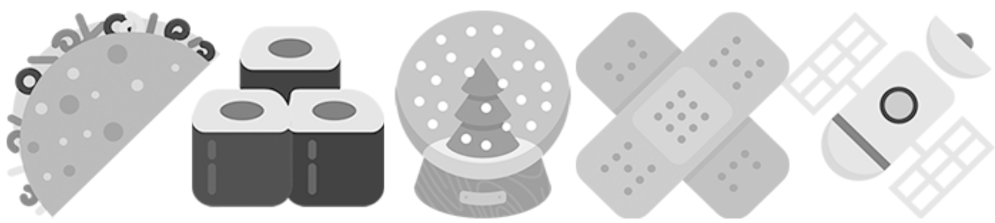

Figure 5. Five of the 20 icons used in the movie clip conditions (Left to right: taco, sushi, snow globe, band aid, and satellite), downloaded from the icon database https://www.flaticon.com.

Third, clips and icons were selected and combined into one sequence of 20 clips with each clip followed by one of the 20 objects in one of the four conditions. The sequence of clips was arranged so that clips from the same movie were placed as far away from each other as possible on the timeline. Each clip lasted roughly 30 seconds prior to the gaze cue. The starting point was made at the start of a scene or the start of a sentence that is logical to the viewer. As a result, some clips differed a few seconds in length. Four versions were made of this final sequence so that each clip resembled a different condition in every version. For example, in version one, the clip of Batman showed an object (car) in the congruent condition for 200ms. Other versions showed the Batman clip in a different condition with a different object. The objects used in every version can be found in Appendix B.

3.4 Design

A 2x2 within-subjects experimental design was used. The independent variables used were Congruency with the gaze (congruent vs. incongruent) and Display time of the object (200ms vs. 800ms). Each condition received an even number of clips. Therefore, each condition received five movie clips from the total for 20 movie clips. As mentioned above, each version of a clip contained a different object. Furthermore, the combination of clips and icons was randomized using a semi-randomized structure, taking into account possible identifiable relationships. An overview of how all objects have been divided over the conditions and clips can be found in Appendices B and C.

3.5 Instrumentation

In Qualtrics, a distraction task and a recognition task were made. This short distraction task also served the purpose of gaining additional participant data to check for outliers. Questions in the distraction task regarded age, gender, the frequency of movie/series viewing, their familiarity with the movies shown in the different clips, and if they wore lenses or glasses during the experiment. Additionally, the familiarity question in the distraction task aimed to detect participants that had already seen the clips in the experiment. Then, the recognition task. This task consisted of a checkboard collection of objects, 40 in total. Half of these objects, 20, were used in the experiment. The participant's goal was to indicate which objects they saw during the viewing task. By clicking on an object, the participant was able to select and highlight objects that he remembered seeing. A screenshot of this event can be seen in Figure 5.

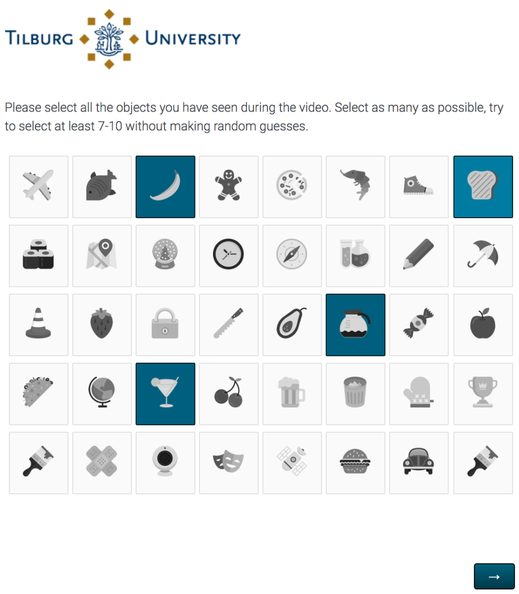

Figure 5. A visual example of the recognition tasks made in Qualtrics. Four highlighted objects are selected of a total of 40 options.

3.6 Procedure

The experiment was conducted in the DCI Lab on Tilburg University. Upon entrance, participants received an information letter and a consent form. Both of these documents are included in Appendix D and E. After giving consent, the participants were positioned in one of the experimental booths where the eye-tracking devices were located. After seating themselves, participants were asked to sit in a comfortable position that allowed them to sit still for the entire experiment. Once the participant was seated correctly, the experimenter positioned the monitor, provided the participant with headphones, and executed a 9-point calibration for the eye-tracker. Finally, the participant received general instructions with regards to the experiment and was asked to read the remaining instructions on the screen. These instructions described that several clips would be shown they needed to pay close attention to them. All given instructions (spoken and on-screen) can be found in Appendix F and G. If there were no further questions, the participants could start the experiment after the experimenter left the booth and closed the door. The experiment began with the viewing of 20 movie clips with a gaze cue followed by objects in one of the four conditions. More specifically, every one of the 20 clips consisted of two seconds of the white cross, a movie clip of approximately 30 seconds, and an object in the congruent or incongruent condition shown for 200ms or 800ms. The white cross was added in between the object and the start of the next clip to make sure that participants returned their gaze towards the center of the screen. This sequence of cross – movie – object was repeated 20 times for one participant. An example of this sequence of clips can be seen in Figure 6.

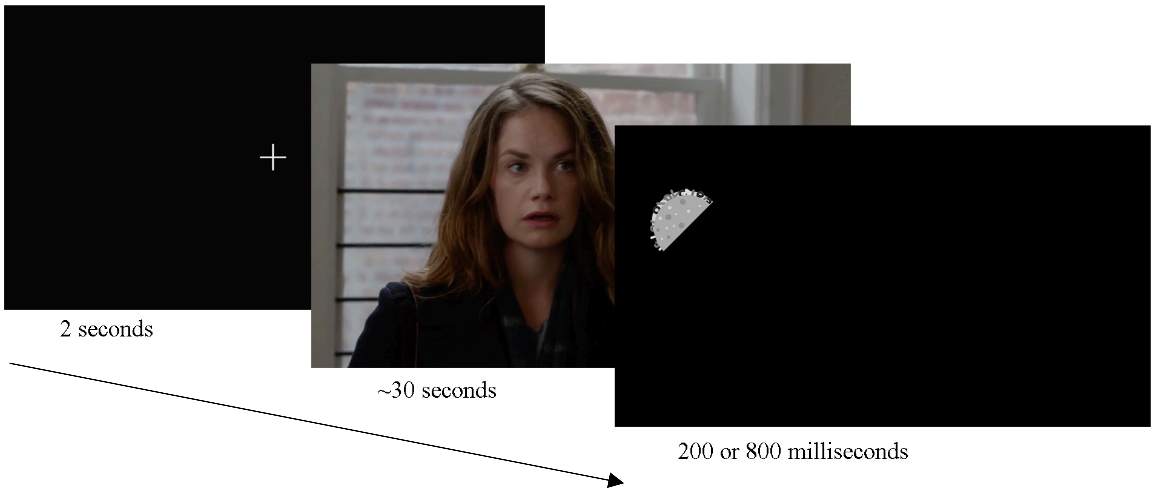

Figure 6. One trial of the experiment. This trail is repeated 20 times for each participant with 20 different movie clips showing all four conditions.

After completing the viewing task, participants had to knock on the door to alert the experimenter to set up the second part of the experiment. When the experimenter left the room again, the participants were allowed to continue with the distraction task that consisted of demographical questions (i.e., age, gender) and control questions (i.e., if they wear glasses, how many movies they watch per week, how many movies they recognized from the clips). The final part of the experiment was the recognition task that started upon completion of the distraction task. Here, participants had to select the objects that they recognized from the clips they saw. They were asked to select a minimum of seven objects without making guesses (participants could continue with the task if they selected less). Completing this task showed the participant the end screen of the survey that included a thank you note and contact information from the researcher and supervisor. After leaving the booth, the participants received a short debrief and were given the opportunity to ask any questions that they might have and the option to leave their email to receive the results of the study. The full experiment lasted approximately 30 minutes.

3.7 Data Analysis

The results of the experiment consisted of eye-tracking data and Qualtrics data. Eye movements were analyzed with the program fixation. To do so, the original movie clips needed to be segmented into parts that needed to be analyzed (i.e., the gaze cue and object) and parts that did not need to be analyzed (i.e., the rest of the movie clip). By changing the timeline in Premiere Pro to frames, the onset and offset times of the segments could be registered. For each segment, a still was made of the gaze cue that represented the corresponding clip. For the onset of each gaze cue segment, the frame in which the gaze cue was initiated was noted. The frame count of five frames (~200ms) after the object disappeared was noted as the offset of each 800ms gaze cue segment. Ten frames (~400ms) after the object disappeared were used for the 200ms condition to include. The additional frames were added to include potential trial fixations in the analysis. Frames counts were converted into milliseconds. This resulted in a list of time frames consisting of onsets and offsets in milliseconds. This final result was used as input for SMImovie. This program used this list to segment the participant's eye-tracking data according to the onsets and offsets. This resulted in only the essential segments of eye-tracking data to be available for further evaluation.

Screenshots were made of each scene consisting of a gaze cue and its successive object. These two were combined in Photoshop into a single screenshot showing both the gaze cue and the object in its original position (see Figure 7). Subsequently, Areas of Interest (AoI) were created. The first area of interest for each clip was fitted around the face of the person initiating the gaze cue. The second AoI was fitted around the location of the object. An example of these AoI's can be seen in Figure 7.

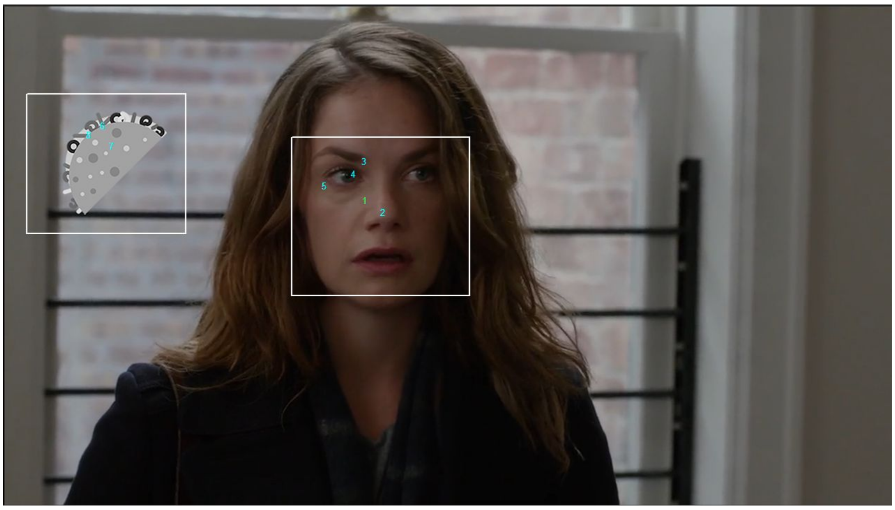

Figure 7. Screenshot of the process of assigning AoIs to faces and objects in Fixation. The first AoI is always the face of the individual in the clip and the second AoI is the object that follows after the gaze cue. The numbers represent fixations during that time frame.

The analysis was performed with Fixation (Cozijn, 2006). This program automatically assigns fixations to the corresponding AoIs. However, if the fixation did not fall within an AoI, it had to be determined whether it should be assigned manually. Taking into account the duration and size of the saccade, the duration of the fixation, and the direction of the saccade, fixations were assigned to AoIs. On some occasions, fixations that fell outside the AoI or in between both showed no good reason to be manually assigned to either of those AoIs. In those cases, fixations were left unassigned and were not included in the analysis. For example, if, in Figure 7, fixation number three would have been positioned on her forehead near her hairline, it would not be assigned to the AoI of the face. However, as this is still near the AoI and there are no other AoIs closeby, it may be assumed that this fixation is mainly due to eye tracker in-accuracy rather than real fixations. Therefore, it can manually be assigned to the AoI of the face. The sum of all the assigned fixations computed the viewing time per AoI per participant. The result of the analysis was imported into SPSS for statistical analysis.

At the end of the experiment, participants were asked to complete a recognition task to see how many objects they memorized. Participants were asked to write down their participant number in the distraction task, and this was later used to retrieve the version of the experiment they saw in SPSS. For each version, a list of the correct answers was made. For each participant, their total number of answers given (The number of correctly highlighted images), correct answers, and correct answers within each condition were computed. The dependent variable Recognition Score was the number of correct answers given. Additionally, variables were made for correct answers in each of the four conditions (200ms Gaze, 800ms Gaze, 200ms NoGaze, 800ms NoGaze).

Results

As previously mentioned in section 2.1 (Participants), In total, out of the 49 participants, the data of 39 participants were included in the analysis. Seven participants were excluded due to having bad data, and three participants did not pass the calibration test for the eye tracker. The data of the remaining 39 participants were analyzed using an analysis of variance of two independent variables (congruency vs. display time) with two conditions (congruent vs. incongruent and 200ms vs. 800ms).

Furthermore, each eye-tracking data file was prepared by including only fixations on the object's AoI. In total, 2569 fixations were made in between the onset of the gaze and offset of the object. 1651 (64.3%) fixations were made on the individual in the movie initiating the gaze cue, 267 (10.4%) of the fixations were outside of both AoIs, and 651 (25.3%) fixations were on the object's AoI (343 on congruent objects, and 308 on incongruent objects). These 651 fixations were used for further analysis.

4.1 Recognition analysis

Recognition scores of the same 39 participants were derived from the data. A Two-way ANOVA in a within-subjects analysis was performed. The independent variables were Congruency (congruent vs. incongruent) and Display time (200ms vs. 800ms.). Out of the total of 780 potentially recognizable objects across all conditions, 374 (48%) objects were recognized. Of which, more specifically, 200 (25,6%) belonged to the congruent condition, and 174 (22,3%) belonged to the incongruent condition. Using the questions in the distraction task, some participants that scored high on familiarity were inspected in more detail. A higher familiarity with the movies showed in the clips did not significantly influence the participants' recognition score. An overview of the means of the recognition scores can be found in Table 2.

Table 2

Mean recognition scores and standard deviations as a function of Congruency (Congruent and Incongruent) and Display time (200ms and 800ms)

	Congruent		Incongruent	
Display time	*M*	*(SD)*	*M*	*(SD)*
200ms	2.07	(1.09)	2.11	(1.10)
800ms	3.61	(1.26)	3.14	(1.21)

There was no effect of Congruency within participants: $F_1(1,27) = 1.81$, $p = .190$. However, in the within-item analysis, an effect was found: $F_2(1,16) = 6.51$, $p = .021$, $\eta^2 = .289$. The findings of the within-participant analysis of Congruency show a pattern that is in line with expectations. More importantly, findings of the within-item analysis support H1: Objects shown in the congruent location cued by the gaze cue will have higher recognition scores compared to objects in the incongruent location.

There was an effect for Display time. Recognition scores for objects displayed for 800ms were higher than recognition scores for objects displayed for 200ms. This was supported by the results: $F_1(1,27) = 34.51$, $p = <.005$, $\eta^2 = .561$; $F_2(1,16) = 51.25$, $p = <.005$, $\eta^2 = .762$. Objects shown for 800ms had higher recognition scores compared to objects shown for 200ms.

There was no interaction between Congruency and Display time: $F_1(1,27) = 2.17$, $p = .152$; $F_2(1,16) = 0.07$, $p = .802$. However, an effect for the 800ms condition was found (Although marginal for the within-item analysis): $F_1(1,38) = 5.47$, $p = .025$, $\eta^2 = .126$; $F_2(1,19) = 3.88$, $p = .064$, $\eta^2 = .169$. No effect was found for the 200ms condition: $F_1(1,27) = 0.02$, $p = .885$; $F_2(1,16) = 2.79$, $p = .114$. Results showed that when the display time was 800ms, recognition scores for objects in the congruent location were significantly higher than recognition scores for objects shown in the incongruent location. This result supports H1. A visual representation of this interaction is provided in Figure 8.

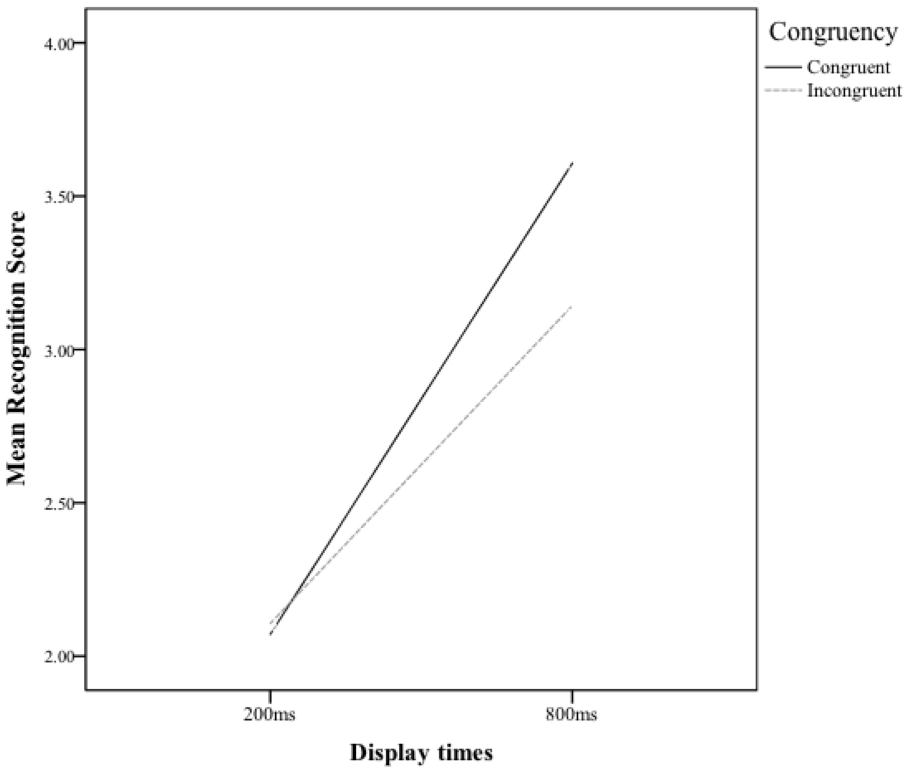

Figure 8. Mean recognition scores as a function of Congruency (Congruent and Incongruent) and Display time (200ms and 800ms).

4.2 Eye-movement analysis

To test the hypothesis whether objects in the congruent location cued by the gaze cue would be viewed longer than objects cued in the incongruent location, a two-way ANOVA was performed. The independent variables were Congruency (congruent vs. incongruent) and Display time (200ms vs. 800ms.) in a between-subjects analysis (due to data loss when analyzed as within-participants in SPSS). An overview of the means of Congruency and Display time can be seen in Table 3.

Table 3

Mean viewing time (ms) and standard deviations as a function of Congruency (Congruent and Incongruent) and Display time (200ms and 800ms)

	Congruent		Incongruent	
Display time	*M*	*(SD)*	*M*	*(SD)*
200ms	305	(43.15)	378	(49.58)
800ms	471	(40.67)	424	(40.26)

An analysis was done between participants F_1 and between items F_2. There was no effect of Congruency: $F_1(1,117) = 0.09$, $p = .771$; $F_2(1,75) = 0.24$, $p = .630$. Within the 800ms condition, results were in line with H2, but not significant.

An effect was found of Display time. Results indicated an effect in both the between participants analysis and the between items analysis: $F_1(1,117) = 5.93$, $p = .016$, $\eta^2 = .048$; $F_2(1,75) = 14.33$, $p = <.005$, $\eta^2 = .160$. Objects that were displayed for 800ms were fixated significantly longer than objects displayed for 200ms.

There was no interaction: $F_1(1,117) = 1.91$, $p = .170$ and $F_2(1,75) = 1.32$, $p = .254$. Additionally, no effects were found for both 200ms and 800ms of Display times: 200ms $F_1(1,50) = 1.15$, $p = .288$; $F_2(1,37) = 0.16$, $p = .689$ and 800ms $F_1(1,67) = 0.73$, $p = .397$; $F_2(1,38) = 2.03$, $p = .163$. These results did not support H2. The data are visualized in Figure 9. As can be seen in Figure 9, for a display time of 800ms, results aligned with H2 and with the findings of the recognition task.

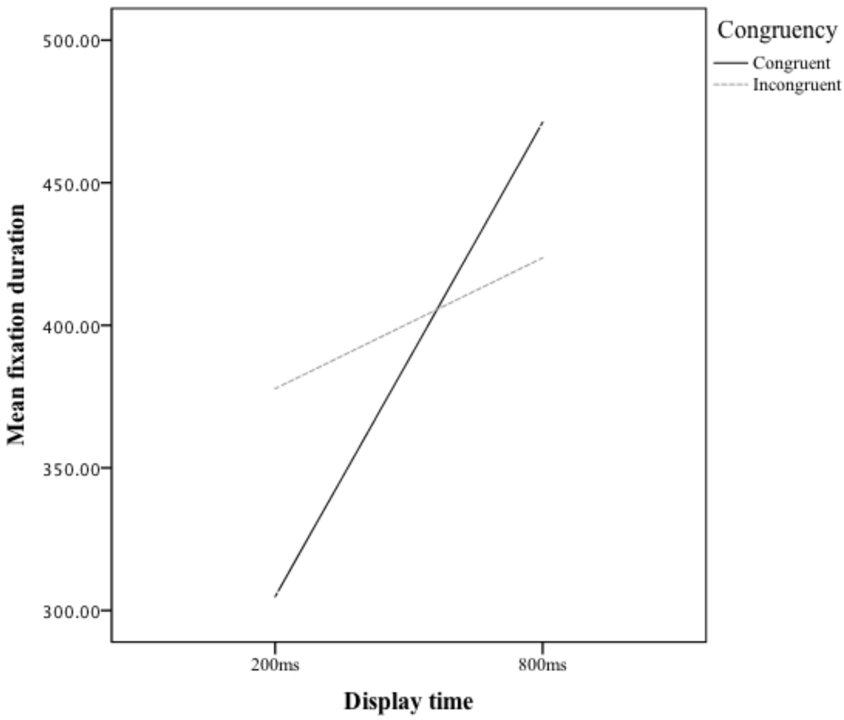

Figure 9. Mean fixation durations (ms) as a function of Congruency (Congruent and Incongruent) and Display time (200ms and 800ms).

Discussion

The present study aimed to investigate how movies direct visual attention, by focusing on one of the attentional cues, as stated by Smith (2006). To do so, an experiment was conducted in which participants' eye movements were recorded whilst they performed a viewing task and a recognition task (without eye-tracker). This was done to measure viewing times and recognition scores for objects showed after the gaze cue with two factors: Placement of the object compared to the area induced by the gaze cue (congruent vs. incongruent) and the duration of display of the object (200ms vs. 800ms).

For the recognition task, the results of this study showed that Congruency significantly influenced participants' recognition in the item analysis. Objects showed in the congruent location were remembered significantly better than objects showed in the incongruent location. Additionally, as expected, results indicated that longer display times significantly increased recognition scores. Lastly, no general interactions were found. Yet within in the 800ms condition of Display time, Congruent objects were remembered significantly better in both the participant and the item analysis. All mean differences between conditions were in line with H1, except for the 200ms condition. A difference in mean values was found in the participant and item analysis for the 200ms condition. Results in the participant analysis showed that, when displayed for 200ms, incongruent object were remembered better. This was the opposite for the item analysis. To conclude, these results indicated that there is indeed a significant influence of the gaze cue on the participants' attention, and, as a result, his recognition. H1, objects shown in the congruent location cued by the gaze cue would have higher recognition scores compared to objects in the incongruent location, has partially been confirmed. Due to mixed findings for Congruency in the 200ms condition. These results only support H1 when objects are displayed for 800ms.

For the viewing task, results indicated that Congruency did not significantly influence the time participants viewed objects. Overall, objects shown in the congruent location were not viewed significantly longer than objects shown in the incongruent location as cued by the gaze cue. However, as expected, an effect for Display time was found. Objects displayed for a longer time were viewed significantly longer than objects displayed for a shorter time. Lastly, no interaction effect was found. Similar to the recognition task, only the means of Congruency for a display time of 200ms showed adverse outcomes relative to H2. the 200ms condition aimed to show the effect of covert attention, as explained by Posner (1980). It was assumed that participants were only able to view objects in the 200ms condition if they were guided by covert attention directed by the gaze. However, results were not significant for Congruency in the 200ms condition, this study fails to fully support H2: Objects in the congruent location cued by the gaze cue will be viewed longer than objects cued in the incongruent location. In conclusion, due to the absence of significant results, H2 is not supported. However, the mean viewing times in the 800ms did show promising results. Differences in viewing times for Congruency when shown for a Display time of 800ms did show findings that were in line with the results of the recognition task and H2.

The results of this study did not fully align with the stated hypotheses. Multiple possible explanations could be given for these findings. These explanations aim to answer why no effect was found for a display time of 200ms, why the means for this condition even oppose the hypothesis, and why no the eye-tracking results for Congruency returned significant but were in line with the hypothesis for a display time of 800ms.

Firstly, the display time of 200ms might be too short to notice explicitly. In an experimental study by Fischer and Ramsperger (1984), they investigated the occurrence of express saccades in human saccadic behavior and reaction times. These saccades are faster than regular saccades and mostly happen when the individual is on higher alert based on previous stimuli (In the experiment this is the disappearing of the previous image with a 200ms gap so participants know that they can expect the target at any time). Participants were asked to move their eyes from a central point to a peripheral visual stimulus as fast as possible. Results showed that the express saccades needed 100ms to reach the target. Participants had to make a saccade of roughly four degrees. In the present study, the angle of the saccade to an object had a maximum value of 15 degrees to the right or left. As saccades can easily reach the speeds of more than 400 degrees per second, the chances seem very low that it would have had an substantial influence (Ventre, Vighetto, Bailly, & Prablanc, 1991). Another study by Ramsøy and Overgaard (2004) explored how minimal a stimulus could be before it would be perceived. Their results showed that for individuals to recognize shapes and positions with about 80% certainty, they need 120-155ms. Taking into account both of these results and the Display times of the current study, it may be reasonable to doubt the appropriateness of choosing 200ms for the short condition.

Secondly, a potential explanation for these outcomes might be the on-screen distance between the gaze cue and the object. As mentioned in the previous statement, 200ms seems too short of a display time to jump to and perceive the object. Therefore, one might increase the time to solve it. Another solution would be to decrease the distance between the object and the gaze cue. Experiments on covert and overt attention, for example, Posner's (1980) overt attention experimental model, used a setup where the distance between the left and right target stimuli were identical. However, in this study, the difference between these distances seemed to be larger. To test this assumption, the horizontal distances (in pixels) between the gaze cue (the point between the eyes) and the object (center) were collected from all clips in all versions. Objects shown in the congruent condition (M = 551.23, SD = 108.49) were further away than objects shown in the incongruent condition (M = 345.82, SD = 79.72). This difference was significant: $t(78) = 7.70$, $p <= .005$. Hence, it can be assumed that overall viewing times and recognition scores for congruent objects in the 200ms condition were influenced by the extra distance that participants had to travel to perceive the object at hand. In the same way, it can be assumed that this difference in distance influenced the 800ms condition.

A third potential explanation is that the gaze cue without head rotation might be too subtle to induce effective attentional guidance. This study chose to mainly use eye-movement-only gaze cues. It was assumed that this would possibly be the cleanest execution of a gaze cue. The chances are that these cues are less powerful because they are too subtle. Pileliene and Grigaliunaite (2018) investigated multiple different cues and their influence on visual attention. However, it should be noted that they used static advertisement. Nevertheless, results showed that participants took less time to make their first fixation on the advertised product when guided by a pointing cue compared to a gaze cue. A possible explanation for this might be the larger size of the cue as the pointing cue, in general, uses more visual space (i.e., an arm) to carry out the cue compared to the gaze cue (i.e., only eyes). Therefore, attracting attention faster.

The study done by van de Schepop (2018) used the match-action cue as the attentional cue to test in his experiment. He was able to identify effects in both the 200ms and 800ms conditions. Although the current study is a partial replication of the study by van de Schepop (2018), the attentional cues used in both studies differ primarily on one aspect. That is the type of cue. The match action cue is an attentional attraction cue. This means that the power of the cue is based on attracting attention by movement. Additionally, these cues use the speed and direction of the movement to form a visual blur that hides the cut. Contrastingly, the gaze cue is an attentional guidance cue. These cues are used to create a perceptual question in the viewer's mind in order to match these expectations after the cut (i.e., making the viewer think "what is he looking at" when the actor gazes in a certain direction before the cut). Attentional guidance cues seem to be less intrusive in general and may also be potentially less powerful in movies compared to attentional attraction cues.

In the present study, three of the 20 movie clips consisted of a gaze with (partial) head rotation. A small statistical test was done using these three clips to test whether using clips with head rotation would show a different outcome. Results indicated that there was no significant effect found for Congruency: $F_1(1,50) = 1.01$, $p = .321$, and Display time: $F_1(1,50) = 0.04$, $p = .845$. Please note that these results are based on only three items and may not be representative. Future research may further investigate the difference in effect between gaze cues with and without head rotation.

Implications

Within science, this study aimed to contribute to the general knowledge of cognitive science. Although not all results were in line with the hypotheses, they do support theories like the *Attentional Theory of Cinematic Continuity* by Smith (2012) by showing support for a priori continuity. Similarly, this study contributes to the knowledge of continuity editing and therefore broadens the knowledge on how, and why, filmmakers use these cues. Alternatively, filmmakers could use these results to better indicate when, and how, to apply the cue in unfamiliar cases. Lastly, adding to the work of van de Schepop (2018), this study contributes to research about another attentional cue to the collection of research on these cues.

In the advertising domain, there are multiple instances where the results of this study might support marketers. For example, one could pitch for commercials including a gaze cue. Companies would then use the shot after the gaze cue for product placements. The marketer is then able to propose an area where the product should be shown, even charging an additional fee for using that "most wanted spot". Additionally, this study strengthens the findings of the previously mentioned advertising studies including gaze cues. More specifically, as research on gaze cues in advertising was mainly located within the web and print domain, this research on gaze cues in the video domain could help broaden the field.

Limitations and future research

Every study has its limitations, and this study is no exception. The cause of most of these limitations has been thoroughly explained in the previous sections. This section, however, will append two limitations with suggestions for future research.

Limitations

First, the 200ms condition for display time might have been too short. Future research may do a more in-depth pretesting phase with different display times or even add a third display time that hovers between the 200ms and 800ms. Hereby it would be possible to distinguish effects that could occur somewhere between 200ms and 800ms. Using other display times such as 350ms and 900ms might still be sufficient for indicating the effect of covert and overt attention.

The second limitation of this study was the difference in on-screen distance between the objects in each condition and the gaze. The significant difference that was found could have had a considerable influence on the results. A lower distance for the congruent objects could have increased the viewing times and recognition scores the same way a higher distance for the incongruent objects could have decreased the viewing times and recognition scores in that condition. Therefore, future research should aim to keep these on-screen distances alike. However, finding clips with these exact characteristics might be a tough challenge, so the use of custom-made shots should also be an option worth considering.

Future research

Alongside with examples of future research based on these limitations, this study proposes some other more general examples of future research. Both the match-action cue and the gaze cue have been explored by van de Schepop (2018) and this study. The study by Pileliene and Grigaliunaite (2018) showed a stronger effect of the pointing gesture cue for initial fixation times. Therefore, it might be interesting to explore the influence of the pointing gesture cue on visual attention in the video domain. Results of such a study could, at the same time, benefit Smith's (2012) work on the AToCC by providing more knowledge about each cue. On the same note, future research on other cues might eventually enable the bundling of all these studies on individual cues into one collaborative work.

Lastly, as mentioned in the limitations, the potential difference in the effect of the gaze cue with and without head rotation has yet to be explored. Therefore, this may be a fruitful topic of investigation for future research. Additionally, the making of these clips could be executed in a lab environment where the same location, props, and actors are used to minimize other effects and solely focus on the gaze cue.

References

Anderson, J. D. (1998). *The reality of illusion: An ecological approach to cognitive film theory* (1th ed.). Southern Illinois University Press.

Bazin, A. (1967). *"The ontology of the photographic image",* in *What is Cinema?* Trans. Hugh Gray, Berkeley. Los Angeles and London: University of California Press, 9–16.

Berliner, T., & Cohen, D. (2011). The Illusion of Continuity: Active Perception and the Classical Editing System. *Journal of Film and Video, 63*(1), 44-63. doi:10.5406/jfilmvideo.63.1.0044

Bordwell, D., Staiger, J., & Thompson, K. (1985). *The classical Hollywood cinema: Film style & mode of production to 1960* (1th ed.). Columbia University Press.

Carroll, N. (1985). The Power of Movies. *Daedalus, 114*(4), 79–103.

Carroll, N., & Seeley, W. P. (2013). Cognitivism, psychology, and neuroscience: Movies as attentional engines. In A. P. Shimamura (1th ed.), *Psychocinematics: Exploring cognition at the movies,* New York, NY, US: Oxford University Press. 53-75. doi:10.1093/acprof:oso/9780199862139.003.0003

Cozijn, R. (2006). Fixation [Computer software]. Tilburg University, Tilburg, Netherlands.

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*(1), 191–195.

Fischer, L. (2007). Film Editing. A Companion to Film Theory (eds T. Miller and R. Stam), 65–83. Blackwell Publishing Ltd. doi:10.1002/9780470998410.ch5

Gibson, J. (2015). The Ecological Approach to Visual Perception. *New York: Psychology Press.* doi:10.4324/9781315740218

Gregory, J. R. (1961). Some psychological aspects of motion picture montage. *Dissertation Abstracts, 22*(1), 317. (n.p.).

Hildreth, E. C., & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience*, *10*(1), 477–533. doi:10.1146/annurev.ne.10.030187.002401

Hutton, S. B., & Nolte, S. (2011). The effect of gaze cues on attention to print advertisements. *Applied Cognitive Psychology*, *25*(6), 887–892. doi:10.1002/acp.1763

Itakura, S. (2004). Gaze-following and joint visual attention in nonhuman animals. *Japanese Psychological Research*, *46*(3), 216–226. doi:10.1111/j.1468-5584.2004.00253.x

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. doi:10.1016/0010-0285(92)90007-O

King, G. (2002). *New Hollywood cinema: An introduction.* London: I. B. Tauris.

Kracauer, S. (1997). *Theory of film: The redemption of physical reality*. Princeton University Press.

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6–7), 925–944. doi:10.1080/13506280902826775

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological Bulletin*, *81*(12), 899. doi:10.1037/h0037368

Pileliene, L., & Grigaliunaite, V. (2018). Effect of visual cues on static advertisement viewing patterns. *Economic Science for Rural Development Conference Proceedings*, (48).

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi:10.1080/00335558008248231

Pudovkin, V. I., & Montagu, I. G. S. (1958). *Film technique and film acting*. Vision.

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23. doi:10.1023/B:PHEN.0000041900.30172.e8

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461

Sajjacholapunt, P., & Ball, L. J. (2014). The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. *Frontiers in Psychology*, *5*, 166. doi:10.3389/fpsyg.2014.00166

Shimamura, A. P. (2013). Psychocinematics: Issues and directions. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies*, New York: Oxford University Press, 1-26. doi:10.1093/acprof:oso/9780199862139.003.0001

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science*, *14*(1), 44–48. doi:10.1111/j.0963-7214.2005.00332.x

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267. doi:10.1016/S1364-6613(97)01080-2

Smith, T. J. (2006). *An attentional theory of continuity editing* (Doctoral dissertation). University of Edinburgh. Edinburgh, United-Kingdom.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*(1), 1–27. doi:10.3167/proj.2012.060102

Smith, T. J. (2013). Watching you watch movies: Using eye-tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies*. New York: Oxford University Press. 165–190 doi:10.1093/acprof:oso/9780199862139.003.0009 van de Schepop, E. (2018). *Ready, set, match action! The influence of a match-action cut on visual attention* (unpublished master's thesis). Tilburg University, Tilburg, Netherlands.

Ventre, J., Vighetto, A., Bailly, G., & Prablanc, C. (1991). Saccade metrics in multiple sclerosis: Versional velocity disconjugacy as the best clue? *Journal of the Neurological Sciences, 102*(2), 144–149. doi:10.1016/0022-510x(91)90062-c

Sekuler, R. (1996). Motion Perception: A Modern View of Wertheimer's 1912 Monograph. *Perception, 25*(10), 1243–1258. doi:10.1068/p251243

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience*, *4*, 5.

Wertheimer, M. (1912). Experimentelle studien uber das sehen von bewegung. *Zeitschrift Fur Psychologie*, *61*, 161–265.

Appendices

Appendix A

List of movies and series used in the experiment and original timing of the gaze cue

Clip	Movie	Season / Episode	Onset Cue	Description	Source
1	Batman: The Dark Knight	N/A	00:19:50	Gaze towards the right	https://www.youtube.com/watch?v=dUMUH4NNYDY
2	Peaky Blinders	S2:EP6	00:06:55	Gaze towards top left	https://www.netflix.com/watch/80002479?source=35
3	Suits	S1:EP9	00:38:50	Looks up, Gaze towards the right	https://www.netflix.com/watch/70195800?source=35
4	The Accountant	N/A	01:51:25	Gaze towards the right	https://www.youtube.com/watch?v=PVBEY9BYPWU
5	The Affair	S1:EP9	00:10:06	Gaze towards the right	https://www.netflix.com/watch/80027745?source=35
6	The Bankjob	N/A	00:06:15	Gaze towards top left	https://www.youtube.com/watch?v=igBIQQV_RI0
7	The Sinner	S2:EP1	00:35:41	Gaze towards to top right	https://www.netflix.com/watch/80175802?source=35
8	Eminem: 8 Mile	N/A	00:14:23	Gaze towards to top right	https://www.youtube.com/watch?v=IdpHxmRFMT0
9	Highwaymen	N/A	01:41:11	Gaze towards the top right / right	https://www.netflix.com/watch/80200571?source=35

10	Suits	S1:EP9	00:10:43	Gaze towards the top right / right	https://www.netflix.com/watch/70195800?source=35
11	Sully	N/A	00:21:06	Gaze towards the right	https://www.youtube.com/watch?v=X3jxqqCLFUM
12	The Accountant	N/A	00:42:17	Gaze towards the left	https://www.youtube.com/watch?v=PVBEY9BYPWU
13	The Affair	S1:EP1	00:05:02	Gaze towards the left	https://www.netflix.com/watch/80027745?source=35
14	Peaky Blinders	S1:EP1	00:06:10	Gaze towards bottom right	https://www.netflix.com/watch/80002479?source=35
15	The Bankjob	N/A	01:02:42	Gaze towards the right	https://www.youtube.com/watch?v=igBIQQV_RI0
16	Southpaw	N/A	00:38:43	Gaze towards the right	https://www.youtube.com/watch?v=Z5z25ypjASA
17	Johnny English: Reborn	N/A	01:28:09	Gaze towards to top right	https://www.netflix.com/watch/70184054?source=35
18	The Fast and The Furious 1	N/A	01:37:46	Gaze towards the right	https://www.youtube.com/watch?v=JXK0LCEXcok
19	The Bankjob	N/A	01:28:03	Gaze towards the top left	https://www.youtube.com/watch?v=igBIQQV_RI0
20	The Affair	S2:EP5	00:03:41	Gaze towards to top right	https://www.netflix.com/watch/80027745?source=35

Appendix B

List of objects used with each clip for each version of the experiment

Clip	Movie	V1	V2	V3	V4
1	Batman: The Dark Knight	Umbrella	Webcam	Apple	pint
2	Peaky Blinders	Cone	satellite	band-aid	snow globe
3	Suits	Candy	Car	Theatre	bread
4	The Accountant	apple	band-aid	pint	cone
5	The Affair	taco	Globe	sushi	satellite
6	The Bankjob	theatre	clock	fish	sushi
7	The Sinner	fish	sushi	satellite	band-aid
8	Eminem: 8 Mile	satellite	theatre	snow globe	taco
9	Highwaymen	band-aid	gingerbread	oven glove	oven glove
10	Suits	gingerbread	pint	candy	umbrella
11	Sully	pint	pizza	clock	gingerbread
12	The Accountant	oven glove	fish	umbrella	apple
13	The Affair	bread	snow globe	globe	fish
14	Peaky Blinders	snow globe	oven glove	bread	pizza
15	The Bankjob	sushi	cone	taco	globe
16	Southpaw	pizza	taco	cone	candy
17	Johnny English: Reborn	globe	bread	car	car
18	The Fast and The Furious 1	webcam	apple	pizza	theatre
19	The Bankjob	clock	candy	gingerbread	clock
20	The Affair	car	umbrella	webcam	Webcam

Appendix C

The distribution of experimental conditions for each version of the experiment

A = Congruent, B = Incongruent | 1 = 200ms, 2 = 800ms. (i.e., A2 = Congruent condition for 200ms)

Clip	Movie	V1	V2	V3	V4
1	Batman: The Dark Knight	B2	A1	A2	B1
2	Peaky Blinders	A1	A2	B1	B2
3	Suits	A2	B1	B2	A1
4	The Accountant	B1	B2	A1	A2
5	The Affair	B2	A1	A2	B1
6	The Bankjob	A2	B1	B2	A1
7	The Sinner	B1	B2	A1	A2
8	Eminem: 8 Mile	B2	A1	A2	B1
9	Highwaymen	A1	A2	B1	B2
10	Suits	B1	B2	A1	A2
11	Sully	A2	B1	B2	A1
12	The Accountant	A1	A2	B1	B2
13	The Affair	A2	B1	B2	A1
14	Peaky Blinders	A1	A2	B1	B2
15	The Bankjob	B1	B2	A1	A2
16	Southpaw	A1	A2	B1	B2
17	Johnny English: Reborn	B1	B2	A1	A2
18	The Fast and The Furious 1	B2	A1	A2	B1
19	The Bankjob	A2	B1	B2	A1
20	The Affair	B2	A1	A2	B1

Appendix D

The information letter used in the experiment

Information letter

Study Name	Researcher	Principal Investigator
The processing of movies	Jules Eekelaar j.a.c.eekelaar@uvt.nl	Reinier Cozijn r.cozijn@uvt.nl +31134662937

This information letter provides all information you need before participating in this experimental study. Please read the information in this letter carefully and contact the researcher if you have questions or need more information to cooperate in this study. The results will be used in a Master's thesis. This study aims to enhance general knowledge about how viewers process and understand movies.
During the experiment, you will be asked to view several movie clips and pay close attention to them. Questions about these movie clips will be asked afterwards. The experiment will last no longer than 30 minutes.

Procedure & risks
This study has been approved by the Ethical Review Board of Tilburg School of Humanities and Digital Sciences. There are no physical or psychological risks involved. There are no right or wrong answers; all data is valuable. Your eye movements will be recorded with an eye tracker. The eye tracker measures eye movements unobtrusively with infrared light. Light emission that meets this standard is not harmful to the human eye.
Your participation in this study is voluntary and we will take the utmost care to treat your personal details confidentially. The only personal details that are being asked are your age, gender, education, movie interests, and whether you wear glasses/lenses. If you decide to take part in this study, you are still free to withdraw at any time, without providing any reason, with no costs (aside from the 0.5 HSP credit that is only given upon completion). If you withdraw from the study before or after data collection is completed, all data will be destroyed.
If you do take part, your eye-tracking data will be entered anonymously into a dataset that will be stored for a period of ten years. These data will not be linked to your personally. After ten years, all data will be destroyed. Data will only be shared between the researcher and the principle investigator mentioned above. Both are fully aware that the data are confidential and should be treated with respect. If we publically present the data, it will only be results of the analysis and no personal details will be disclosed.

Compensation
An 0.5 point in Human Subject Pool credit will be given upon completion of the experiment as compensation for participation in this study. Additionally, participants will receive proof of participation in scientific research of Tilburg University, signed by the principal investigator.

Contact information
If you have questions after this study, or you experience adverse effects as a result of participating in this study, please feel free to contact the researcher or principal investigator whose contact information is provided above.

In the following informed consent, you will officially be requested to participate in this study. If you decide to take part, we really appreciate your participation in our research and want to thank you in advance!

Appendix E

The informed consent letter that participants signed to participate in the experiment

Informed consent

Informed Consent Form for Participants
Tilburg University
Department of Communication and Cognition

Study Name	**Researcher**	**Principal Investigator**
The processing of movies	Jules Eekelaar j.a.c.eekelaar@uvt.nl	Reinier Cozijn r.cozijn@uvt.nl +31134662937

Please read this document carefully. Your signature is required for participation. You must be at least 18 years of age to give your consent to participate in research.

The information letter provided all the necessary information to decide to take part in this study. If you have any further questions about the study, the information letter or the informed consent please ask them now.

In this informed consent, we would like to ask you to confirm the following statements:

I have read and I understand the provided information of the information letter and have had the opportunity to ask questions. I am at least 18 years old. I understand that my participation in this study is voluntary and that I am free to withdraw at any time, without giving a reason and without costs. I agree that my eye movements are monitored. I agree that the research data will be stored for a period of ten years and that the data I provided will be processed without being linked to my personal details. I agree that the results of this study may be used for a scientific publication. I understand that I will be given a copy of this consent form. I voluntarily agree to take part in this study.

______________________________ __________________
Participant’s Signature Date

____
Participant’s Name

______________________________ __________________
Researcher’s Signature Date

____

Appendix F

The verbal instructions given prior to the start of the experiment

Instruction

You are about to take part in a movie experiment. Please read the instructions carefully, and if you have questions do not hesitate to ask me.

You will see several movie clips and all you need to do is pay close attention while watching them. Afterwards, you will be asked questions about them. During the experiment, your eye movements will be monitored. You will not notice that while watching. To use eye-movement recording, the eye tracker needs to be calibrated. I will explain to you shortly how that works and what it is that you need to do. For eye tracking to succeed, it is important that you sit still during the experiment and try not to move too much with you head.

Once calibrated, you will start with the experiment and I will leave the booth. You will then see several movie clips in a row. Before each clip, a screen appears with a white cross in the middle of the screen. Please look at the cross until it disappears and then watch the movie. This procedure is repeated for each clip.

Is everything clear? Do you have any questions? Okay, let’s begin.

Appendix G

On-screen text of the experiment

On-screen text

Screen 1 – before the start of the calibration

"We are now going to do a calibration for the eye-tracker. Instructions will follow. The calibration can be done multiple times so don't be afraid to fail.
You can press the space bar once to go to the calibration screen."

Screen 2 – After calibration before watching the movie clips

"During the experiment, you will be asked to view several different movie clips. Please pay close attention to them.
Questions about these movie clips will be asked afterwards. If you have no further questions, you may start as soon as I have left the room and have closed the door.
Try to focus on the white cross between clips to continue to the next clip.

Good luck!

To start the experiment, press the spacebar."

Screen 3 – After watching the clips, before the second part of the experiment (Qualtrics)

"The first part of the experiment is done. You can knock on the door to alert the experimenter"