The Attentional Theory of Cinematic Continuity

Tim J. Smith

Birkbeck, University of London

2011

Word count, body text = 9,318 words

Word count, total = **11,601** words

Corresponding Author:

Tim J. Smith

Psychology, Malet Street, Birkbeck, University of London

London, UK, WC1E 7HX

tj.smith@bbk.ac.uk

Phone: +44 (0)20 7631 6359

Fax: +44 (0)20 7631 6312

Biographic information:

Dr. Tim J. Smith is a lecturer in the Department of Psychological Sciences, Birkbeck, University of London. His research covers all aspects of Visual Cognition with special focus on the active perception of naturalistic scenes via eye movements. His personal and professional passion for film has led to his applying his empirical methods to the issue of how we perceive film.

Abstract

The intention of most film editing is to create the impression of continuity by editing together discontinuous viewpoints. The techniques used to achieve this, the continuity editing rules are well established yet there exists an incomplete understanding of their cognitive foundations. In this essay I will present the *Attentional Theory of Cinematic Continuity (AToCC)*. AToCC identifies the critical role visual attention plays in the perception of continuity across cuts and demonstrates how perceptual expectations can be matched across cuts without the need for a coherent representation of the depicted space. The theory explains several key elements of the continuity editing style including match-action, matched-exit/entrances, shot/reverse-shot, the 180° rule and point-of-view editing. AToCC formalizes insights about viewer cognition that have been latent in the filmmaking community for nearly a century and demonstrates how much vision science in general can learn from film.

Introduction

> *"Looking at real things, the human vision fastens itself upon a quick succession of small comprehensible incidents, and we form our eventual impressions, like a mosaic, out of such detail… **The director counterfeits the operation of the eye with his lens.**"* (D. W. Griffith, 1926, quoted in Jesionowski, 1982; page 46)

The majority of film, television and dynamic visual media are composed according to a suite of rules developed in the United States in the first 20 years of the 20th century: *the Continuity Style* (or Hollywood Style; Bordwell, Staiger, & Thompson, 1985). This style is implemented via a suite of heuristics that assist a filmmaker in deciding how to shoot a scene so it will cut together for "continuity" in the editing suite. These heuristics are collectively known as the *Continuity Editing Rules (CE Rules).* Over 90% of shots surveyed from Hollywood film adhere to the CE rules (Bordwell, et al., 1985; Levin & Wang, 2009). These conventions are open to interpretation, experimentation, and extension (e.g. Bordwell, 2006) but they remain as the foundation of how all moving-images are edited. However, nobody knows what "continuity" is. Continuity is not the editing style but the intended outcome of the style. A film edited for continuity is often said to make the viewing process effortless and the editing "invisible" (Dmytryk, 1986). *"The purpose of the [continuity editing] system is to tell a story coherently and clearly, to map out the chain of characters' actions in an undistracting way. Thus editing, supported by specific strategies of cinematography and mise-en-scene, is used to ensure narrative continuity...... to create a smooth flow from shot to shot*" (Bordwell & Thompson, 2001 pg. 284). Statements such as this are common in the film literature but rarely endeavor to go further towards an actual definition of continuity. What exactly is flowing from one shot to another? Every cut is by definition a discontinuity of the whole image. How is it possible that one type of cut results in continuity but another doesn't when in terms of the raw image they are both essentially the same?

In this essay I will argue that in order for us to understand continuity editing we need to understand the role of the viewer in the perceptual construction of a film. The "flow" refers to the continuity of viewer cognition: what they are attending to, what they are perceiving, and what they are expecting. There is no continuity without a viewer. I will outline a theoretical framework within which we can discuss the collaborative act of continuity construction between the film and the viewer. The *Attentional Theory of Cinematic Continuity (AToCC)* acknowledges that the viewer is active, even when sat stationary in a cinema auditorium, and through their gaze they seek out information on the screen, formulate expectations about future events, attend to objects across cuts, and represent minimal details of a scene that are relevant to the narrative. The continuity editing rules use natural attentional cues such as off-screen sounds, conversational turns, motion, gaze cues, and pointing gestures to trigger attentional shifts across cuts. The combination of attentional cues pre-cut, and matching minimal expectations post-cut allow viewer cognition to precede seamlessly from shot to shot, scene to scene, sequence to sequence, and across the entire narrative.

I will begin by outlining current theories of continuity and will argue that a fully adequate theory does not yet exist. Applying insights from empirical studies I will formalize *AToCC* and use it to make predictions about when and how continuity will be perceived across a selection of cuts. Eye movements recorded during film viewing will provide a measure of continuity. The gaze behavior across a selection of cut types will show 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.

Hollywood Style and the Ecological View of Film Perception

The continuity editing rules are widely known and discussed in detail elsewhere (e.g. Arjon, 1976; Bordwell & Thompson, 2001; Dmytryk, 1986; Katz, 1991). As such I will only provide a brief overview of them here.

A typical scene presented in the Hollywood Style will involve the *analytical breakdown* of a scene. This entails the presentation of a scene as a series of shots filmed increasingly closer to the main characters. A typical scene might begin with an establishing shot depicting the layout of the scene and the spatial relationship of the key characters. As the action progresses the camera will gradually move closer to the characters either presenting multiple characters on the screen at the same time from the waist up (Medium Shot), one character viewed over the shoulder of another character (*over the shoulder* shot), or a single character in close-up (a *single*). Shots will typically alternate between the characters engaged in the interaction (*shot/reverse-shot*). If two shots of the same character or object are cut together the viewpoints must be at least 30° away from each other or entail significantly different shot sizes (the 30° Rule; violation of which creates a "*jump cut*"[1]). As characters move out of shot the camera will cut back to a long shot to reveal the new spatial arrangement (*re-establishing* shot). If a character leaves the screen (*matched-exit/entrance*) or moves within a shot (*match-on-action*) a cut can be made to the continuation of their motion. Glances off camera can be followed by the target of their gaze in order to show what they are looking at (*point-of-view* shot). The filmmaker is free to use any viewpoint of the action as long as the camera never crosses the *axis of action*, an imaginary line drawn between the main characters of the scene. For example, once a scene is established from one side of the line joining two conversing characters all subsequent shots must be filmed from within the 180° arc around this line.

[1] I will not have space to describe the cognitive motivation for the 30° rule here. Other theorists have identified the of origin of the "jump" as unintended apparent motion (Anderson, 1996; Hochberg & Brooks, 1978).

Adhering to this 180° Rule is believed to be important as crossing the line reverses the screen direction of the action and where characters are looking relative to the screen (i.e. their eye lines).

If you watch any TV program, Film, Cartoon, or read a comic book you will be able to identify this pattern of shots. These conventions emerged through trial-and-error, self-experimentation by filmmakers, experimentation on audiences with box office as the dependent measure, and the rapid transmission (i.e. copying) between filmmakers and across countries and the globe. But why did cinema settle on this compositional form so early on in its history?

One proposed explanation of why the continuity editing works is that it produces a stream of audiovisual information that is similar enough to our experience of the real-world to allow us to perceive it directly, without any specific cognitive skills (Anderson, 1996; Bordwell, et al., 1985; Cutting, 2005; Gibson, 1979; Lindgren, 1948). This *ecological* view of film cognition was best summarized by Joseph Anderson in his book *The Reality of Illusion* (1996): "*the perception and comprehension of motion pictures is regarded as a subset of perception and comprehension in general, and the workings of the perceptual system and the mind of the spectator are viewed in the context of their evolutionary development."* (pg. 10). Certain cinematic techniques may have developed as pure convention, such as fades and dissolves to represent the passage of time (Bordwell, 1996), but the majority of the techniques are thought to have settled in their final form due to their compatibility with our existing cognitive abilities (Anderson, 1996).

However, simply stating that film is "good enough" for our visual system to "solve" ignores the gross differences between film and reality. The exact form of these differences has been examined in depth by other theorists (Arnheim, 1957; Cutting, 2005; Hochberg & Brooks, 1978b) but their essence can be summarized as 1) Film can instantaneously change the visual scene across a cut; 2) Film can represent scenes and events in a piecemeal fashion, by juxtaposing views of objects that are not spatially, or even temporally adjacent; 3) Redundant sections of actions, periods of time, or extents of space can be elided, i.e. removed.

A theory of how we perceive cinematic continuity must specify either how we overcome these incompatibilities or why they are unimportant. Current ecological theories (Anderson, 1996; Cutting, 2005; Gibson, 1979) provide a starting point for such a theory but do not apply sufficient understanding of how we perceive dynamic real-world scenes in order for clear predictions to be made about how the CE Rules optimize the compatibility with existing cognitive abilities. To begin developing such a theory I will first turn to the oft ignored task of trying to define "continuity".

What is Continuity?

definition: *continuity = 1. logical sequence, cohesion, or connection; 2. a continuous or connected whole; 3. the comprehensive script or scenario of detail and movement in film or broadcast.* (Collins English Dictionary, 1998)

According to the dictionary definition, something can be said to have "continuity" if it is continuous or part of a whole. On the surface, this would seem to be a strange thing to attribute to edited moving-images. If a filmmakers objective was to successfully represent continuity of a scene the easiest option would be to film it in one continuous shot as was the tradition in early film and later in films that utilized long-takes and deep-focus such as Orson Welles' *Citizen Kane* (1941). In fact, this argument was put forward by film theorist André Bazin (1967) when he criticized the continuity editing style for restricting a viewer's experience of a scene. However, in a long-shot of a scene objects of interest occupy a very small portion of our visual field and due to visual acuity limitations of the human eye we are unable to make out the fine detail. Filmmakers discovered very early on that cutting in to close-ups provides greater information about details of a scene and frees film from the restrictions typically faced by theatrical productions.

But deciding to cut together shots introduces the problem of making them seem continuous. How do you create the impression that a scene is continuous when it is constructed from discontinuous units? Most books on film theory and practice devote large sections to solving the problem without first defining what continuity is. For instance, Daniel Arjon's wonderfully encyclopaedic *Grammar of the Film Language* (1976) provides 624 pages of solutions for creating continuity in a wide variety of scenes without once defining continuity. Across the literature, continuity is variously defined as creating the illusion of continuous action (Dmytryk, 1986; Katz, 1991; Reisz & Millar, 1953; Salt, 2009), preserving graphic, space, time, logical, and narrative connections between shots (Anderson, 1996; Bordwell & Thompson, 2001; Dmytryk, 1986; Katz, 1991; Reisz & Millar, 1953; Salt, 2009), the continued presence of at least one actor (Salt, 2009) or background landmarks across a shot (Hochberg & Brooks, 1978b), avoiding a noticeable "jerk" (Dmytryk, 1986; Reisz & Millar, 1953), drawing attention to the cut (Bordwell & Thompson, 2001) or details of the film's production (Sandys, 2005). What is apparent from this list is the multiplicity of functions the concept of "continuity" is believed to fulfill. One theme that is common to most discussions of "continuity" is the creation of a spatiotemporally continuous scene. For example, if a scene entails two people engaged in conversation the scene will be edited together in such a way that the viewer will perceive the action as playing out in real-time and in a coherent space even though this perception may span a selection of shots, viewpoints, and may distort the passage of time.

However, most contemporary books on film editing question whether the continuity editing rules should be strictly adhered to and instead encourage filmmakers to experiment with and extend the continuity style. Evidence of this stylistic change can be found in the changing frequency with which cuts violate the continuity rules. In 1959, jump cuts were virtually non-existent in Hollywood films but by 1999 most films contain a handful of jump cuts (Salt, 2009). This de-emphasis of strict adherence to the continuity editing rules has been supported by the editor, Walter Murch who placed the preservation of three-dimensional continuity of a scene at the bottom of a list of factors for an editor to consider: 1) Emotion, 2) Narrative, 3) Rhythm, 4) Eye trace, i.e. where the viewer is looking, 5) Planarity, i.e. 3D transposed to 2D, and 6) 3D continuity (Murch, 2001).

In summary, it appears that when most filmmakers and film theorists refer to "continuity" they mean coherent space and continuous time. However, they also acknowledge that "continuity" is an illusion created in the mind of a viewer and not an inherent feature of the stimulus itself, i.e. the film. As such, a theory of cinematic continuity must focus on the relationship between the film and the viewer. To gain an insight into how we may perceive continuity in film our first recourse should be to look at the perception of continuity in the real-world.

Perceiving continuity in the real-world

The issue of continuity in film is obvious as film is inherently discontinuous. However, in a real-world scene questioning how we perceive continuity seems like an odd question because the laws of physics denote that the real-world must have continuity. Objects cannot change, move or disappear without going through certain, physically constrained transformations (Gibson, 1979). If an object suddenly appeared or disappeared, the associated visual transients should capture attention and alert a viewer to the discontinuity. As such viewers can assume the world is continuous unless the world itself alerts us to a discontinuity

(Dennett, 1991; O'Regan, 1992). Let's call this assumption of continuity *a priori continuity,* i.e. independent of evidence.

However, our visual system is imperfect. We can only see about 120° of a scene at any one time and over 90% of this is in very low resolution due to the uneven distribution of photosensitive cells on the back of the eye (the *retina*). As we cannot encode detail from the whole visual field simultaneously we need to rotate our eyes so that the light from the area of interest is projected on to the most sensitive central part of the retina. These eye movements are easily observed if you watch another person's eyes as they read a page of text. You will notice that their eyes make a series of quick jumps separated by brief moments when they are still. The movements are *saccades* and the stationary moments are *fixations*. We only take in visual information during fixations as a process known as *saccadic suppression* stops us seeing the blurred image streaking across the retina as we move our eyes (Matin, 1974). You can see the result of this process yourself by looking in a mirror as you saccade from one eye to another. If you watch somebody else do it you can see their eyes in flight but if you look at your own eyes you will only see the fixations.

In order for us to perceive a continuous scene, the discrete samples of information encoded during each fixation must be perceptually "stitched" together and added to a working memory representation of the scene. Several theories of scene perception exist and they vary in the degree to which they believe scene information is retained across eye movements (Henderson & Hollingworth, 2003; Hollingworth, 2006; O'Regan, 1992; O'Regan & Noë, 2001; Rensink, 2000). However, for our purposes it is sufficient to note that some detailed information is retained across a saccade and used to form a coherent representation of objects of interest within a spatially consistent scene.

Each saccadic eye movement takes 20-50ms on average and each fixation will last about 200-500ms (Rayner, 1998). This rapid shifts of attention should mean that the visual world has little opportunity to change significantly across a saccadic eye movement and, therefore an assumption of *a priori* continuity should be valid. However, if a scene is artificially changed during a saccadic eye movement, this assumption of continuity may result in viewers failing to notice striking changes to the scene. This phenomenon is known as *change blindness*.

If the visual transients associated with a change are masked, either by coinciding the change with a saccadic eye movement, distracting attention, or obscuring the screen during the change (e.g. with a flicker) viewers may fail to notice changes to details within the scene (see Simons & Levin, 1997; for review). If the viewer fails to detect the change after the distraction it is seen as evidence that they have either not attended to the object, not encoded the details in working memory, failed to retrieve the detail from memory, or not performed *correspondence*: the checking of the visual features they see before them and what is in their memory (Kahneman, Treisman, & Gibbs, 1992). The likelihood of detecting a change is in direct proportion to the amount of attention paid to the object. If the change is close to objects of interest such as a person's head it is more likely to be detected (Rensink, ORegan, & Clark, 1997). Our memory for a scene is populated only by the details that are attended and the longer the time since a detail was last attended, the less likely that detail is to still be active in working memory (Hollingworth & Henderson, 2002).

Most change blindness studies have investigated how we construct spatially coherent representations of static scenes. However, when watching film the viewer also needs to monitor dynamic features such as the movements of characters within a scene. Up until recently dynamic scenes have received very little empirical enquiry (see Mital, Smith, Hill, & Henderson, 2011 for review). A series of striking demonstrations of change blindness in real-world and edited dynamic scenes by Dan Levin and Dan Simons provide the main reference for any theory attempting to explain cinematic continuity (Levin & Simons, 1997; Simons & Levin, 1998). In one experiment, Levin & Simons deliberately introduced continuity errors into a film sequence of two women conversing across a dinner table (1997). The short clip contained nine continuity errors, such as a scarf or plate that changed color across a cut. During initial free-viewing of the scene none of the viewers reported seeing anything odd. Even during a second viewing in which they were instructed to detect changes, viewers only detected an average of 2 out of the 9 changes and tended to notice changes closest to the actors' faces such as the scarf. Although, Levin and Simons did not record viewer eye movements it is clear from my own eyetracking studies (Mital, et al., 2011; Smith & Henderson, 2008) that viewers will mostly be focused on faces and spend virtually no time on peripheral details. Without fixating a peripheral object such as the plate, the viewer is unable to represent the color of the plate in memory and can, therefore not detect the change in color even if they later fixate it.

Our inability to detect continuity errors in film under normal viewing conditions demonstrates that not all details of a scene are attended or represented in enough detail for us to detect a discontinuity. However, even attending to an object does not guarantee change detection (Levin & Simons, 1997; Smith, Lamont, & Henderson, under review). In another study, Levin & Simons (1997) used a matched-exit/entrance to show one actor leaving a room in one shot and then a different actor continuing the action in the next shot. Even though the two actors were wearing different clothes and were of different ages only 33% of viewers reported the change when asked if they had "noticed anything odd". This demonstration of change blindness in a dynamic scene emphasizes the role location and movement of a character (i.e. spatiotemporal information) plays in convincing us that a scene is continuous.

Levin & Simons used this apparent primacy of spatiotemporal continuity in dynamic scene perception to begin formulating a theory of cinematic continuity (2000). They referenced evidence from developmental psychology that spatiotemporal information may provide the foundation for all subsequent object perception. Infants below 10 months of age have been shown to prioritize spatiotemporal information above all other information when perceiving objects. Infants do not display shock when they are shown a toy duck that moves behind an occluder only to reappear on the other side as a ball (Xu & Carey, 1996). Infants older than 10 months will be shocked if the occluder is removed and only one object is revealed. The spatiotemporal continuity of the action as it moves behind the occluder is used to continue perceiving the object even when it cannot be seen. When a different object appears that continues the expected action we initially perceive it as the same object. Only by performing correspondence between our stored representation of object identity and the current form of the object will we detect the change.

Levin & Simons (2000) suggested that this emphasis on spatiotemporal expectations for the perception of object and scene continuity may be critical to understand how we perceive continuity in film. In film, as in the real-world "*people's actions and motives are central, whereas visual detail is only interesting when it clarifies our understanding of these actions and motives.....Whatever impressions we have of continuously occupying the same space and looking at the same scene comes from the consistency of far fewer details than intuitions suggest."* (Levin & Simon, 2000; pg. 376).

Levin & Simons' interpretation of their empirical evidence provides a suitable foundation for a theory of cinematic continuity but they do not develop their theory further enough for it to make predictions about how and when viewers will perceive continuity during film viewing. I will now endeavor to do so.

The Attentional Theory of Cinematic Continuity

Taking my lead from the ecological view of film perception (Anderson, 1996; Cutting, 2005; Gibson, 1979) and combining it with the empirical observation of limited awareness and representation of visual detail during film viewing (Levin, 2010; Levin & Simons, 2000; Levin & Wang, 2009). I will propose an *Attentional Theory of Cinematic Continuity* (AToCC). I will briefly outline the theory before explaining its components and predictions in depth.

The key assumption of AToCC is that viewers do not and should not construct a detailed spatiotemporal representation of the depicted scenes. Such effortful cognition is redundant for the perception of most important elements of a cinematic narrative. Editing a scene in a way that allows the perception of "continuity" is not about enabling the construction of a detailed spatiotemporal representation. Instead it is about enabling the viewer to shift their attention to the audiovisual details currently relevant to them and the narrative. The emphasis of this theory is on viewer attention: what audiovisual feature is the viewer currently attending, how are they shifting attention between features and what expectations do they have about the future form of the attended features? During a shot, a viewer has an *a priori* assumption of continuity for all audiovisual features. However, their attention will be focused on a small, spatially localized subset of these features. If a cut occurs that the viewer is not expecting the visual transients created by the change from one shot to another will capture attention and a discontinuity will be perceived. The continuity editing rules minimize this by cuing viewer attention across the cut. This is accomplished via natural attentional cues such as off-screen sounds, conversational turns, onsets of motion, gaze cues, pointing gestures, departures from our field of view and establishing rhythmical and logical patterns of attention. These cues encourage the viewer to either voluntarily or involuntarily shift their attention and/or gaze to a new part of the scene. Coinciding a cut with this shift minimizes viewer awareness of the visual transients and, therefore the discontinuity created by the cut. Different types of attentional cues establish different types of expectation about what the viewer will see once they reorient to the target of their attentional shift. For example, if an unknown off-screen character begins speaking the viewer will covertly shift their attention in the direction of the speech and have a vague expectation of the visual appearance of the person based on the character of their voice. In general these expectations are minimal and prioritize basic object information such as location on the screen, orientation and direction of action. If these expectations are met within a few frames following a cut, viewer attention will seamlessly shift to the target object and the assumption of *a priori* continuity will be confirmed. If these expectations are not met, or the cut occurs without first cuing attention the viewer will shift into active reconstruction mode, surveying the shot for audiovisual details that may allow them to understand the relationship between the new shot and the previous. If information is located that points to spatiotemporal continuity, e.g. the continuation of a conversation or line of action from the previous shot, *a posteriori* continuity will be perceived (i.e. requiring evidence). This is a more effortful form of continuity perception but is essential for successful comprehension of the depicted narrative. *A posteriori* continuity breaks the flow of attention from one shot to another and, therefore breaks the illusion that the film represents a naturalistic scene. Instead, the viewer has greater awareness of the artificial and constructed nature of the film.

This overview of AToCC should not sound controversial to anybody who has studied editing in-depth or struggled repeatedly to find the right time to cut. My intention is to outline a cognitive theory that both film theorists and filmmakers can understand and apply to their considerations of editing. However, while accessible AToCC is also grounded on current understanding about dynamic scene perception. I will now begin outlining the theoretical foundations for AToCC and its practical implementation.

I will discuss the three stages of AToCC:

1. Attending to a shot
2. Cuing attention across a cut
3. Matching expectations after a cut

Attending to a shot

On initial presentation of a shot we process vague *gist* in parallel across the whole shot (Biederman, 1972). This gist contains basic layout information of the shot, the scene category, and the location objects of potential interest. In order to begin constructing a perceptual representation of a shot we need to serially attend to objects and associate visual details with points in space. We can only maintain 3-4 active object representations in working memory at any one time (Kahneman, et al., 1992) and these representations are bound to a location in the scene by an attentional pointer that allows us to track object movement (Pylyshyn, 1989). Visual features such as appearance, orientation, size, and identity are bound to this attentional pointer within an *object file* (Kahneman, et al., 1992). Without focused attention, objects files are believed to rapidly decay, lose binding between features, and eventually disappear from working memory (Kahneman, et al., 1992). Therefore, knowing where viewers attend in a scene is critical for understanding which objects may be active in working memory and tracked for continuity.

When considering where viewers attend in a shot we need to make a distinction between *covert* and *overt* attention (Posner & Cohen, 1984). Covert attention is the increase of visual processing resources allocated to a point in space *without the eyes moving to the location of the selected feature* (see Findlay & Gilchrist, 2003 for review). Colloquially this is the equivalent of watching something out of the corner of your eye. Overtly attending to a point in space involves coupling attention and the location of gaze so that what is attended is also fixated. Covert attention can only be measured by looking for a benefit in processing the attended location (e.g. quicker reaction times) where as overt attention can be easily measured by recording the location of the eyes using an eye tracker (see Holmqvist et al., 2011; for review of eyetracking methods and technology). Covert attention is often thought to move independent of overt attention. However, studies in which participants are free to move their eyes have shown that covert attention shifts of attention are rare except preceding a saccade (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995). As a result, a record of where a person has fixated will also be a good measure of where they attended (Henderson, 1992).

<<Insert Figure 1 about here >>

Where we fixate is a consequence of two interacting sources of information: *bottom-up*, low-level visual information such as luminance, color, edges, and motion, and *top-down*, higher-order cognitive factors such as our individual preferences, viewing task, and recognition of object and scene semantics. The decision about where we attend is a combination of bottom-up factors such as motion trying to involuntarily capture attention and top-down factors such as interest trying to keep attention under voluntary control. For example, a long panning shot from *There Will Be Blood* (Anderson, 2007) clearly demonstrates the interaction between bottom-up and top-down influences over viewer gaze and the fluctuation of this relationship over time (Figure 1). By superimposing the gaze location of eleven viewers on to this sequence and monitoring the clustering of their gaze (i.e. the heatmap; Mital et al, 2010) we can see how top-down and bottom-up factors interact to guide their attention[2]. Sudden onsets

[2] A video of viewer gaze behaviour in this scene can be viewed here: http://vimeo.com/19788132. Data was gathered on an SR Research Eyelink 1000 tracker. The film was displayed on a 21 inch monitor and viewing position was stabilised with a chinrest. Clearly this is a highly artificial viewing condition but given the right eyetracking equipment more natural viewing conditions can be achieved.

of motion, such as Paul's hand in the doorway (frame 2) or appearance of faces (frames 3, 5, and 6) lead to attentional synchrony. Bottom-up factors such as the presence of bright lights compete with top-down interest in faces. In frame 3, interest in Paul and his speech stops most viewers looking at the light in the top right corner of the screen. Later in the scene Paul's head drifts out of shot and viewers default back to the screen centre (Mital, et al., 2011) and are then captured by the suddenly unoccluded desk lamp (frame 4). This short camera pan illustrates how viewers must reconcile the bottom-up audiovisual factors fighting over their attention with their own top-down interest in order to decide where to attend.

In the absence of editing, viewer gaze will shift around objects of interest within a shot guided by their own interest, low-level audiovisual features of the scene, and compositional features such as camera movements, lighting, focus, color, etc. A filmmaker will attempt to align these three factors to ensure coordination between where they want the viewers to look and where the viewers want to look. If this marriage is successful it will result in *attentional synchrony*. A clear example of this marriage can be seen in a long-take scene from *There Will Be Blood* (Anderson, 2007). Immediately following the scene depicted in Figure 1, Paul, Daniel, Fletcher Hamilton, and HW (Daniel's son) gather around a map while Paul describes the location of his farm (Figure 2). In the absence of editing, this scene has to rely on the staging of the scene, the actor's dialogue and physical exchanges to guide viewer attention[3].

<<Insert Figure 2 about here >>

By superimposing the gaze location of eleven viewers on to this scene we can observe exactly how it is choreographed by the action (Figure 2). Shifts in conversation are followed by clustering gaze on the speaker's face then gradual shifts back to the listener to gauge their

[3] For an extended analysis of the gaze behaviour in this clip and its relation to staging of the scene see a piece I wrote for David Bordwell's blog: http://www.davidbordwell.net/blog/2011/02/14/watching-you-watch-there-will-be-blood/

response (Figure 2; 2nd row). Head turns and shifts of gaze may result in a viewer following the eyeline to the target of the gaze (Figure 2; 3rd row) (Langton, Watt, & Bruce, 2000). Sudden movements such as a character rising up or lifting their hand may involuntarily attract gaze to the source of movement (Figure 2; 4th row; Mital et al, 2010) and if the movement is caused by a hand forming a pointing gesture viewers may shift from the hand to the target of the point (Figure 2; bottom row) (Langton & Bruce, 2000). These audiovisual events result in a remarkable degree of attentional synchrony and provide time points at which the filmmaker knows where the majority of viewers are looking. Each saccadic eye movement is the physical manifestation of a *perceptual inquiry (Hochberg & Brooks, 1978a)*: "Who is speaking?", "What are they looking at?", "What was that movement?", "What are they pointing at?". Each perceptual inquiry elicits expectations about what the expected answer to the inquiry might be. These expectations may be very minimal and primarily focus on the target object and its location relative to the viewer. For example, a gaze shift will direct attention along the direction of the gesture and to the first object it encounters. If no target is present, attention will covertly shift in the direction of the gesture but the eyes will remain at the source of the gesture. The viewer will be ready to locate the answer to their perceptual inquiry as soon as it appears. Such universal perceptual inquiries and the resulting attentional shifts provide the mechanism by which filmmakers can cue attention across cuts and create the illusion of continuity.

Cuing a Cut and Matching Expectations

In order to minimize viewer awareness of the visual transients associated with a cut and maintain the assumption of *a priori* continuity the filmmaker needs to coincide the cut with an attentional shift[4]. Several types of cuts belonging to the continuity style utilize such cues: matched-exit/entrance, match-action, establishing and shot/reverse-shot, point-of-view shot, and point shot. I will outline how each of these types of cut cue attention, create perceptual inquiries and match minimal expectations after the cut to maintain *a priori* continuity. The gaze behavior of multiple viewers will be presented as a demonstration of how seamlessly attention shifts across cuts.

Matched-Exit/Entrance

<<Insert Figure 3 about here >>

Our field of view of a real-world scene is about 120° but if we wish to make a saccadic eye movement to a target more than 30° away from our current point of fixation the saccade will usually be accompanied by a head rotation in the same direction (Land & Tatler, 2009). For example, if an owl flew left-to-right past our field of view and we wished to saccade to its landing position we would first program an eye movement to the right. As our eyes land our head would catch up by rotating right. In order for our gaze to remain in the same location in space our eyes would have to rotate in the opposite direction (i.e. left). This compensatory eye movement is automatic (a process known as *Vestibular Ocular Reflex*). If the owl continued its flight this pattern of eye and head rotations would repeat: saccade right, rotate head right, eyes drift left.

Now compare this pattern of natural eye movements to the shots used to depict the same event in a scene from *Blade Runner* (Scott, 1982; Figure 3). The owl flies out of shot eliciting a perceptual inquiry, "Where is it going?". Our attention covertly shifts off-screen right but given that our eyes cannot leave the screen they instead rebound and saccade to the screen

[4] I am not arguing that the cut occurs during a period in which attention is absent or suppressed during a blink or saccade as empirical evidence indicates that cuts rarely coincide with such periods (Smith & Henderson, 2008). Instead attention is preparing to shift to another location.

centre (see Mital, et al., 2011 for evidence of this central tendency), mirroring the saccade they would have made in a real scene as compensation for our head rotation. Cut to a shot depicting the continuation of the owl's flight provides an analogue for our normal head rotation and allows us to pick up the owl's flight at our new gaze location. Our attention shifts seamlessly from the side of the screen at which the owl departed, to the screen centre, and back to the owl as it reappears[5]. If the second shot had crossed the axis of action (violating the 180° rule) the screen direction of the owl's flight would have reversed and the second shot would no longer be analogous with our normal head rotation or our covert shift of attention off-screen right.

Matched-exit/entrances are a powerful way of giving the impression of a continuous action spanning spatially adjacent locations by simply matching the expected screen entry location and direction of motion of an object. This effect was elegantly demonstrated by the first empirical film theorists, Lev Kuleshov and Vsevolod Pudovkin (1929). In one experiment, they used matched-exit/entrance cuts to create an impossible space by showing a man and woman walking towards each other in a series of separate shots and then meeting in a final shot. Even though each shot was filmed in a completely different location the screen exits focus viewer attention on character motion and minimize awareness of the discontinuous background. The matched-exit/entrance created the impression of a continuous space but this does not mean that viewers actively maintain a coherent representation of this new combined space. Empirical evidence suggests that if background features are not relevant to the viewing task they will not be attended and therefore not make it into long-term memory (Hollingworth, 2006) or be used to construct a spatial representation (Levin & Wang, 2009).

[5] This pattern of eye movements in response to a matched-exit/entrance was first reported by Edward Dmytryk in his book *On Filmmaking* (1986). Remarkably, Dmytryk's description was based purely on introspection as at that time eyetracking technology was not capable of recording eye movements during film viewing.

Match-Action

<<Insert Figure 4 about here >>

Matched-exit/entrances can be seen as a specific instance of a match-action cut. Sudden onsets of movement within a shot draw attention to the screen location of the movement and its future trajectory (Mital, et al., 2011). Expectations about the visual scene and object features are abandoned as we focus on the spatiotemporal details of the action. If a match-action cut is timed to coincide with the onset of the action and presents a new viewpoint of the same action at the same screen location as pre-cut the viewer's expectations will be satisfied and *a priori* continuity will be perceived.

Figure 4 (left column) presents a match-action cut from *Blade Runner* (Scott, 1982). Deckard's descent towards the bottom of the frame triggers a cut to an over-the-shoulder shot of Rachael in which her face is in the same screen location as Deckard's before the cut, leading to a smooth transference of attention across the cut. The same shift in attention could have been accomplished by a whip pan between Deckard and Rachael, such as that used in Lars Von Trier's *Dancer in the Dark* (2000; Figure 4, right column) but the impact on viewer attention is massively different. In the match-action, viewer attention is smoothly transferred between the characters by being deposited in the right screen location by the motion. During the whip pan, the sudden unexpected horizontal flow of the whole visual field results in uncoordinated gaze as viewers scurry to the right edge of the screen in an attempt to locate the target of the pan. Viewer gaze manifests a pattern of eye movements usually only observed during self-motion: *Optokinetic reflex (OKR)*. This is an involuntary pattern of saccades to moving targets followed by pursuit in the opposite direction. OKR eye movements are typically coordinated with feedback from the vestibular system about self-motion. During a whip pan, the presence of OKR eye movements may result in a mismatch between the head rotation implied by the flowing image and our absence of corresponding vestibular signals. Such a mismatch often results in a feeling similar to seasickness.

In a cut detection task I have shown that match-action cuts are the hardest type of cut to detect (Smith & Henderson, 2008). A third of all match-action cuts were missed when they were presented in clips taken from feature films. The combination of capture of attention by the motion, the sudden obscuring of the screen, and the continued motion following the cut guides attention across the cut and gives the viewer a strong sense of *a priori* continuity.

This power of match-action editing can also be used to create the illusion of continuity even in the face of impossible transformations of the rest of the scene. A film which makes considerable use of this technique is Alan Resnais' *Last Year in Marienbad* (1961). Resnais' film is about the fallibility of memory and creates a fantastical space in which time, character and action is amorphous and dreamlike. In one particular cut (Figure 5; top row) a woman is shown turning around as if something over her shoulder has caught her attention. As she is midway round a match-action cut transports the viewer to a completely different scene as the woman completes her turn. Her motion carries our eyes through the cut and we infer *a priori* continuity. It is only after we have oriented to the new shot that we become aware of the change and perceive an *a posteriori discontinuity*. However, inferring a discontinuity from the presence of incompatible evidence comes too late to alter the initial perception of continuity.

<<Insert Figure 5 about here >>

The use of match-action editing to create impossible transformations was first demonstrated by early filmmaker and stage magician, George Méliès. He discovered that by stopping the camera mid action, changing the scene, and then continuing filming he could make a scene spontaneously transform. The tricks depicted in Figure 5 (bottom row) shows how this technique can be used to make objects, such as a woman hidden under a blanket and a stool disappear during a movement. The continuous background and movement gives the impression of *a priori* continuity but Méliès' skillful violation of our expectations shows this assumption to be violated, resulting in a striking comic effect.

Establishing shot and Shot/reverse-shot sequence

<<Insert Figure 6 about here >>

One of the most recognizable elements of the Hollywood Style is the shot/reverse-shot sequence during dialogues. The location of two or more characters on the screen are initially established by presenting all characters on screen simultaneously (an *establishing shot*). The sequence then alternates between shots favoring each character in turn, typically while they are speaking. The shot/reverse-shots can be over-the-shoulder shots in which the back of the listener is seen out of focus on the opposite side of the screen to the speaker or singles, a close-up depicting only the speaker (Figure 6; left column). Whatever the type of shot the most important considerations are 1) the camera never crosses the axis of action, and 2) eyelines match across a cut, i.e. meet on the screen. These considerations are often thought to allow viewers to construct a coherent spatial representation of the scene (Berliner & Cohen, 2011). However, recent empirical evidence suggests that while viewers can comprehend the spatial layout of shot/reverse-shot sequences (Frith & Robson, 1975; Kraft, 1987; Kraft, Cantor, & Gottdiener, 1991; Levin, 2010) spatial comprehension does not seem to be important during normal viewing as they fail to notice when the line is crossed (d'Ydewalle, Desmet, & Van Rensbergen, 1998; Germeys & d'Ydewalle, 2007). What does seem to be important is rapid orienting to the content of the new shot (Germeys & d'Ydewalle, 2007). This can be achieved by establishing clear screen locations for each character (Block, 2001)[6].

For example, in the *Blade Runner* shot/reverse-shot sequence (Figure 6; left column) the viewer can maintain attentional pointers for Bryant (left) and Deckard (right) relative to where they are expected to appear on the screen and relative to each other. Rather than construct a coherent 3D representation of the space, the viewers simply need to know that Deckard is to the right of Bryant. When Bryant is presented in close-up viewers fixate his face but covertly shift their gaze to screen right in anticipation of Deckard's appearance. As soon as they hear Deckard speak their attention begins shifting to his previous screen location (about 200ms after speech onset). If the next shot presents Deckard in this screen location viewer expectations will be matched and their assumption of *a priori* continuity will be satisfied. Compare this to the sequence from *Requiem for a Dream* (Aronofsky, 2000; Figure 6; right column) in which the first cut to a close-up crosses the line and the framing of each character means that their eyelines don't match across the cut. Violating expectations of screen location may mean that even if a cut is cued by a shift in speaker the viewer may not perceive *a priori* continuity as they have to work against the covert attentional shift caused by the actor's eyeline and instead saccade in the opposite direction to find the speaker.

<<Insert Figure 7 about here >>

Conversational turns and shot/reverse-shot sequences can also be used to create anomalous perception of continuity. In another scene from *Last Year in Marienbad* (Resnais, 1961) a conversation between two men is initially presented in a long medium shot, allowing the

[6] A method for monitoring how viewer attention shifts across cuts was first proposed by Bruce Block in his book *The Visual Story* (2001). Block showed how centres of interest such as faces can be colocated across cuts in order to aid the transference of viewer attention from one shot to another and create continuum of attention. Block based his analyses on introspection about what as likely to be the centre of interest but now, with the help of eyetracking we can see exactly where viewers are looking and how the centre of interest changes over time.

viewer to saccade between both men as they speak. As the conversation shifts to the standing man on the right a cut carries us to a close-up with the man framed in the same screen location and facing the same direction. Viewer attention shifts seamlessly across the cut and *a priori* continuity is perceived. However, as the camera slowly pulls back we realize that the man is now seated and time has elapsed. Our perception of *a posteriori* discontinuity creates a mismatch with the initial perception of continuity leading to a strange sense of unease.

Gaze and Point shots

<<Insert Figure 8 about here >>

Gaze and point cues can also shift attention to off-screen space and create minimal expectations that bridge a cut. As overt attention can never leave the screen, a character's glance off-screen will produce a covert shift in the direction of their glance. If the target of their glance is not known this will create a perceptual inquiry ("What are they looking at?") and an expectation that the inquiry will be answered by an object oriented in a way compatible with the glance. The attentional shift will then be satisfied by the eyes remaining in the same screen location (covert only) or by saccading to the target (covert + overt). Either way, the attentional cue bridges the cut and maintains *a priori continuity* irrespective of other irrelevant discontinuities e.g. object features, scene, space.

For example, in this sequence from *Blade Runner* (Scott, 1982; Figure 8) the spatial relationship between three characters, Rachael, Deckard and the owl are initially established using just gaze cues. The three characters are never shown occupying the same space. The coincidence of glances off-screen with dialogue ensures that viewer attention covertly shifts along each character's eyeline in anticipation of the cut. Viewer attention can either be cued to the opposite part of the screen and the target of the gaze cue presented in this new location (as in the initial cut from Rachael to Deckard; Figure 8). Gaze cues can also be used to covertly shift attention off-screen but keep the viewer eyes in the same screen location by cutting to the target presented in the same screen location. Shots 4, 5, and 6 (Figure 8) accomplish this very well. The cut between the owl (3) and Deckard (4) is slightly off meaning that the owl's gaze shift will cue viewer attention in the opposite direction to the saccade required to shift back to Deckard's face. Such a mismatch may result in a violation of *a priori* continuity. Although the correction saccade required to locate Deckard is so small that it may occur automatically.

Complex cuts and combination of cues

So far I have identified how attentional cues operate in isolation across cuts. However, for a lot of cuts there will be redundancy with multiple cues guiding attention across a cut and establishing the relationships between shots. For example, diegetic sound such as dialogue or environmental sounds often continue across a gaze-match or match-action cut to reinforce the temporal continuity implied by the visual events. Off-screen sounds can also be used to shift covert attention off-screen and cue a cut to the source of the sound.

Repeated patterns of crosscutting or rhythmical editing can establish expectations of when and to what a cut occurs (Cutting, DeLong, & Nothelfer, 2010). For example, in the firing range sequence from *Last year in Marienbad* (Resnais, 1961; Figure 9) we begin the scene by being uncertain what the men are doing (shot 1). As each of the men turns, fires a gun off screen and we see targets hit by the bullets (shot 2) we establish a causal connection between the shots that allow us to perceive *a posteriori* continuity. The repeated alternation of these shots builds up considerable expectations about what will happen when one of the men shoots off-screen. This expectation is used to startle the viewer when the protagonist turns to shoot (shot 3) but instead of cutting to the same shot of the targets we cut to a woman walking in the same room at a different point in time (shot 4). Note that this violation is not operating at the level of space, but at the level of object i.e. the woman is the wrong target for our attentional shift. We are drawing casual inferences about how a firing action in the first shot is followed by a shot of the targets we are not constructing a coherent spatial representation of the two spaces (Berliner & Cohen, 2011).

<<Insert Figure 9 about here >>

Other complex editing patterns can create expectations about the timing and types of cuts and the content of new shots. The key is using attentional cues to either overtly direct viewer attention to an on-screen space or covertly shift attention off-screen. If the subsequent cut satisfies our minimal expectations about the form of the target of attention our assumption of *a priori* continuity will be satisfied. In most situations this continuity will imply that time is continuous. However, this is not always the case as certain attentional cues may direct attention to time points, events and locations not spatiotemporally contiguous with the current scene. For instance, a character's announced departure from a location in order to travel to a new location may cue viewers to expect that the next shot will be in the new location. The continuity style is rich with the acceptable use of ellipses and causally and narratively motivated cuts (Anderson, 1996; Bordwell & Thompson, 2001). However, due to space limits I will not have the opportunity to discuss these factors in-depth here[7]

[7] See the work of Jeff Zacks and Joe Magliano for in-depth discussion of these issues (Zacks & Magliano, in press).

Conclusion

In this essay I have presented an overview of an *Attentional Theory of Cinematic Continuity (AToCC)*. Various attempts to understand the relationship between the continuity editing rules (or Hollywood Style) and the perception of continuity have been obstructed by a belief that viewers must draw inferences about the depicted spaces and construct coherent spatial representations of a scene (Berliner & Cohen, 2011; Frith & Robson, 1975; Hochberg & Brooks, 1978b; Smith, 2006). However, recent evidence of change blindness and memory across saccades indicate that such complete and coherent representations are not maintained during normal vision. Instead, viewers have minimal expectations about how perceptual inquiries will be answered (Hochberg & Brooks, 1978b). These inquiries are manifest as either overt or covert attentional shifts triggered by changes in the audiovisual scene. Such changes lead to a high-degree of coordination of where multiple viewers attend in a dynamic scene and across cuts (attentional synchrony; Mital, et al., 2011). Once scene gist and the location of a small number of objects of interest is identified attention will shift between these objects in line with depicted events such as onsets of motion, gaze cues, audio onsets and conversational turns. Matching the timing of cuts with the onset of such events can guide viewer attention across the cut as long as minimal expectations about target form, screen location and action are preserved after the cut. Such matches create a continuity of attention across a cut and result in smooth and easy comprehension of the depicted scene.

It is important to note that adhering to the continuity editing rules may only be one way to satisfy expectations cued across cuts. Continuity may be perceived in films that deviate, extend or completely abandon the continuity editing style (see Bordwell, 2006 for discussion of how recent film may *intensify* our perception of continuity). For example, music videos, adverts, avant garde and experimental films that do not depict spatially consistent actions or scenes may still be perceived as having continuity depending on how viewer attention is cued prior to a cut and what form their expectations take after a cut.

The intention of this theory is to provide a framework upon which other aspects of film cognition can be built. I acknowledge that many of the aspects deemed key to the cinematic experience such as emotion, characterization, narrative, and style are not currently discussed in AToCC. I am certain that these factors may interact with the issue of continuity perception discussed here. However, for the time being I am happy to leave the discussion of these factors to other theorists more qualified than I on these topics.

In closing, I would like to turn AToCC on its head. The majority of the empirical evidence and theoretical motivation for AToCC presented here has come from recent scientific investigations into visual cognition (including some performed by myself). However, the particular combination of evidence and theory presented here about how we perceive dynamic visual scenes has never previously been attempted. Vision scientists have only recently become interested in how we perceive dynamic scenes and our current understanding is currently rudimentary compared to the insights filmmakers have had for decades. Vision scientists can learn a lot from formalizing and testing the intuitions of filmmakers. It has been nearly a century since a psychologist, Hugo Münsterberg first commented on the potential for understanding real-world perception through film, and over 80 years since Kuleshov and Pudovkin performed their experiments that form the backbone of all film theory, yet even today we only have a rudimentary understanding of the psychology of a pastime that constitutes a sizable portion of our lives. I believe we are now at a turning point. The combination of Cognitive Film Theory, recent developments in Visual Cognition and the advent of new methods for the analysis of how viewers attend to and perceive film and dynamic scenes means that we can make significant advances in our understanding of film cognition.

Acknowledgments

I would like to thank Sam Wass and Paul Taberham for comments on earlier drafts of this paper and my collaborators John Henderson, Parag Mital, Robin Hill and Melissa Võ for their assistance in all things dynamic. The Dynamic Images and Eye Movement (DIEM: http://thediemproject.wordpress.com/) project to which some of this work belongs was funded by the Leverhulme Trust (Ref F/00-158/BZ) . Finally I would like to thank my colleagues in the Society for the Cognitive Studies of the Moving Images (SCSMI) for supporting my work and providing copious guidance and feedback over may conferences. Special thanks to David Bordwell for his continual support and inspiration. Part of this work was developed and described in my PhD thesis (Smith, 2006).

Filmography

Anderson, P. T., 2007, *There Will Be Blood, USA*

Aronofsky, D. 2000, *Requiem for a Dream, USA*

Resnais, A., 1961, *Last year in Marienbad, France*

Méliès, G., 1896, *The Vanishing Lady, France*

Méliès, G., 1898, *The Magician, France*

Scott, R., 1982, *Blade Runner, USA*

Von Trier, L., 2000, *Dancer in the Dark, Denmark*

Welles, O., 1941, *Citizen Kane, USA*

Bibliography

Anderson, Joseph. D. 1996. *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale and Edwardsville, US: Southern. Illinois University Press.

Arjon, D. 1976. *Grammar of the Film Language*. Beverly Hills, CA, US: Simlan-James Press.

Arnheim, R. 1957. *Film as Art. Berkeley*, California: University of California Press.

Bazin, A. 1967. *What is Cinema?* Berkeley and Los Angeles, US: University of California Press.

Berliner, T., and Cohen, D. J. 2011. "The Illusion of Continuity: Active Perception and the Classical Editing System". *Journal of Film and Video*, 631, 44-63.

Biederman, I. 1972. "Perceiving real-world scenes". *Science*, 177, 77-80.

Block, B. 2001. *The Visual Story: Seeing structure of film, TV, and New Media*. MA, USA: Focal Press.

Bordwell, D. 2006. *The Way Hollywood Tells It: Story and Style in Modern Movies*. Los Angeles, US: University California Press.

Bordwell, D., Staiger, J., and Thompson, K. 1985. *The Classical Hollywood Cinema: Flm Style and Mode of Production to 1960*. London, UK: Routledge.

Bordwell, D., and Thompson, K. 2001. *Film Art: An Introduction* Vol. 6th. New York, USA: Mc Graw Hill.

Cutting, J. E. 2005. "Perceiving scenes in film and in the world". Pp. 9-27 in *Moving Image Theory: Ecological considerations,* eds. J. D. Anderson and B. F. Anderson. Carbondale, IL: Southern Illinois University Press.

Cutting, J. E., DeLong, J. E., and Nothelfer, C. E. 2010. "Attention and the evolution of Hollywood film". *Psychological Science*, 21: 440-447.

d'Ydewalle, G., Desmet, G., and Van Rensbergen, J. 1998. "Film Perception: The processing of film cuts". Pp. 357-367 in *Eye guidance in reading and scene perception*, G. Underwood ed.,. Oxford: Elsevier.

Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown.

Deubel, H., and Schneider, W. X. 1996. "Saccade target selection and object recognition: evidence for a common attentional mechanism". *Vision Research*, 36, 1827-1837.

Dmytryk, E. 1986. *On Filmmaking*. London, UK: Focal Press.

Findlay, J. M., and Gilchrist, I. D. 2003. *Active vision: The psychology of looking and seeing*. Oxford: University Press.

Frith, U., and Robson, J. E. 1975. "Perceiving the language of films". *Perception*, 41, 97-103.

Germeys, F., and d'Ydewalle, G. 2007. "The psychology of film: perceiving beyond the cut". *Psychological Research*, 71, 458-466.

Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston, USA: Houghton Mifflin.

Henderson, J. M. 1992. "Visual attention and eye movement control during reading and picture viewing". Pp. 260-283 *in Eye movements and visual cognition: Scene perception and reading*. K. Rayner ed. New York: Springer-Verlag.

Henderson, J. M., and Hollingworth, A. 2003. "Eye movements and visual memory: Detecting changes to saccade targets in scenes". *Perception and Psychophysics*, 651, 58-71.

Hochberg, J., and Brooks, V. 1978a. "Film Cutting and Visual Momentum". pp. 293-317 in *Eye Movements and the Higher Psychological Functions* eds. J. W. Senders, D. F. Fisher and R. A. Monty. Hillsdale, NJ: Lawrence Erlbaum.

Hochberg, J., and Brooks, V. 1978b. "The Perception of Motion Pictures". *Handbook of Perception Vol. X* in eds. E. C. Carterette and M. P. Friedman. New York, US: Academic Press.

Hollingworth, A. 2006. "Visual memory for natural scenes: Evidence from change detection and visual search". *Visual Cognition*, 14, 781-807.

Hollingworth, A., and Henderson, J. M. 2002. "Accurate visual memory for previously attended objects in natural scenes". *Journal of Experimental Psychology-Human Perception and Performance*, 281, 113-136.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. 2011. *Eye Tracking: A comprehensive guide to methods and measures*. Oxford, UK: OUP Press.

Jesionowski, J. E. 1982. *Thinking in Pictures: Dramatic Structure in D. W. Griffith's Biograph films.* Berkeley and Los Angeles: University of California Press.

Kahneman, D., Treisman, A., and Gibbs, B. J. 1992. "The reviewing of object files: Object-specific integration of information". *Cognitive Psychology*, 242, 175-219.

Katz, S. D. 1991. *Film Directing Shot by Shot: visualizing from concept to screen*. CA, US: Michael Wiese Productions.

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. 1995. "The Role of Attention in the Programming of Saccades". *Vision Research*, 3513, 1897-1916.

Kraft, R. N. 1987. "Rules and strategies of visual narratives". *Perceptual and Motor Skills*, 64, 3-14.

Kraft, R. N., Cantor, P., and Gottdiener, C. 1991. "The Coherence of Visual Narratives". *Communication Research*, 185, 601-615.

Land, M. F., and Tatler, B. W. 2009. *Looking and Acting: Vision and eye movements in natural behaviour*. Oxford, UK: OUP Press.

Langton, S. R. H., and Bruce, V. 2000. "You must see the point: automatic processing of cues to the direction of social attention". *Journal of Experimental Psychology: Human Perception and Performance*.252, 747-757.

Langton, S. R. H., Watt, R. J., and Bruce, V. 2000. "Do the eyes have it? Cues to the direction of social attention". *Trends in Cognitive Sciences*, 42, 50-59.

Levin, D. 2010. "Spatial representation of the sets of familiar and unfamiliar television programs". *Media Psychology*, 13, 1-23.

Levin, D., and Simons, D. J. 1997. "Failure to detect changes to attended objects in motion pictures". *Psychonomic Bulletin and Review,* 4, 501-506.

Levin, D., and Simons, D. J. 2000. "Perceiving stability in a changing world: Combining shots and integrating views in motion pictures and the real world". *Media Psychology*, 2, 357-380.

Levin, D., and Wang, C. 2009. "Spatial representation in film". *Projections: The Journal for Movies and Mind*, 3, 24-52.

Lindgren, E. 1948. *The Art of the Film*. London, UK: George Allen and Unwin Limited.

Matin, E. 1974. "Saccadic suppression: A review and an analysis*". Psychological Bulletin*, 8112, 899-917.

Mital, P. K., Smith, T. J., Hill, R. L., and Henderson, J. M. 2011. "Clustering of gaze during dynamic scene viewing is predicted by motion". *Cognitive Computation*, 31, 5-24.

Murch, W. 2001. *In The Blink Of An Eye: a perspective on film editing*. Los Angeles, USA: Silman-James Press.

O'Regan, J. K. 1992. "Solving the 'real' mysteries of visual perception: the world as an outside memory". *Canadian Journal of Psychology*, 46, 461-488.

O'Regan, J. K., and Noë, A. 2001. "A sensorimotor account of vision and visual consciousness". *Behavioral and Brain Sciences*, 24, 939-1011.

Posner, M. I., and Cohen, Y. 1984. "Components of visual orienting". pp. 531-556 in *Attention and performance X* eds. H. Bouma and D. Bouwhuis. London: Erlbaum.

Pudovkin, V. 1929. *On Film Technique*. London: Gollancz.

Pylyshyn, Z. W. 1989. "The role of location indexes in spatial perception: a sketch of the FINST spatial-index model". *Cognition*, 32, 65-97.

Rayner, K. 1998. "Eye movements in reading and information processing: 20 years of research". *Psychological Bulletin*, 1243, 372-422.

Reisz, K., and Millar, G. 1953. *Technique of Film Editing*. London, UK: Focal Press.

Rensink, R. A. 2000. "The dynamic representation of scenes". *Visual Cognition*, 71-3, 17-42.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. 1997. "To see or not to see: The need for attention to perceive changes in scenes". *Psychological Science*, 85, 368-373.

Salt, B. 2009. *Film Style and Technology: History and Analysis* Vol. 3rd. Totton, Hampshire, UK: Starword.

Sandys, J. 2005. *Movie Mistakes: Take 4*. London, UK: Virgin Books.

Simons, D. J., and Levin, D. T. 1997. "Change Blindness". *Trends in Cognitive Sciences, 1*, 261-267.

Smith, T. J. 2006. "An Attentional Theory of Continuity Editing". Ph.D diss.., University of Edinburgh, Edinburgh, UK. http://www.era.lib.ed.ac.uk/handle/1842/1076 (accessed 5th Jan 2012)

Smith, T. J., and Henderson, J. M. 2008. "Edit Blindness: The Relationship Between Attention and Global Change Blindness in Dynamic Scenes". *Journal of Eye Movement Research*, 22, 1-17.

Smith, T. J., Lamont, P., and Henderson, J. M. under review. "The Penny Drops: Change Blindness at Fixation". *Perception*.

Xu, F., and Carey, S. 1996. "Infants' metaphysics: the case of numerical identity". *Cognitive Psychology*, 30, 111-153.

Zacks, J. M., and Magliano, J. P. in press. "Film understanding and cognitive neuroscience". In *Art & the Sense.* eds. D. P. Melcher and F. Bacci. New York: Oxford University Press.

Figure captions

Figure 1: Six frames from a panning shot in *There Will Be Blood* (Anderson, 2007). Paul enters Daniel Plainview's office (frames 1 and 2) and converses with Daniel (off-camera) as the camera pans to follow his movement (frames 3, 4, 5). Fletcher Hamilton is revealed as Paul sits down (frame 6). Gaze locations of eleven viewers are superimposed onto each frame (green circles) and a heatmap representing the degree of clustering of gaze for that frame is also displayed.

Figure 2: Top row: An example frame from a long-take from *There Will Be Blood* (Anderson, 2007). The gaze location of eleven viewers (green spots) are superimposed on to the frame and their coordination is displayed as a heatmap. The hotter the color of the heatmap the more coordinated the gaze. Other rows: The gaze points of multiple viewers is used to create a "peekthrough" heatmap in which each gaze location shines a virtual spotlight on the film frame. Any part of the frame not attended is black and the more viewers look in the same location, the hotter the color. The eight frames show examples of viewer gaze in response to a variety of audiovisual events.

Figure 3: Three shots from *Blade Runner* (Scott, 1982) depicting a matched-exit/entrance of an owl's flight. Gaze location of seven viewers is superimposed over the top if each frame as a heatmap.

Figure 4: Left column: A match-action cut from *Blade Runner* (Scott, 1982). Deckard goes to sit on a chair. As his head nears the bottom of the frame we cut to an over-the-shoulder shot of Rachael with Deckard continuing his downward motion. Right column: A whip pan is used as an alternative to a match-action cut as two characters play catch in *Dancer in the Dark* (Von Trier, 2000). Gaze location of seven viewers is superimposed over the top if each frame as a heatmap.

Figure 5: Match-Action cuts. Top row: Alan Resnais' *Last Year in Marienbad* (1967); Bottom row: George Méliès' *The Vanishing Lady* (1896) and *The Magician* (1898).

Figure 6: Left column: a shot/reverse-shot sequence from *Blade Runner* (Scott, 1982). Right column: a shot/reverse-shot sequence that crosses the line from *Requiem for a Dream* (Aronofsky, 2000). Gaze location of seven viewers is superimposed over the top if each frame as a heatmap.

Figure 7: Erroneous shot/reverse-shot sequence from Alan Resnais' *Last Year in Marienbad* (1967).

Figure 8: A sequence of gaze-cued cuts from *Blade Runner* (1982). Movements by characters within the frame are represented as white arrows. Red arrows represent camera movements. Cuts are represented as red dotted lines. Speech is written beneath the associated shot (D=Deckard, R=Rachael). Gaze location of seven viewers is superimposed over the top if each frame as a heatmap.

Figure 9: The firing range sequence from Alan Resnais' *Last Year in Marienbad* (1967). Repetition of shots of men shooting and the targets being hit creates an expectation of their spatial relationship. Replacing the targets with a woman in the same location violates viewer expectation and creates shock.

Figures

Figure 1

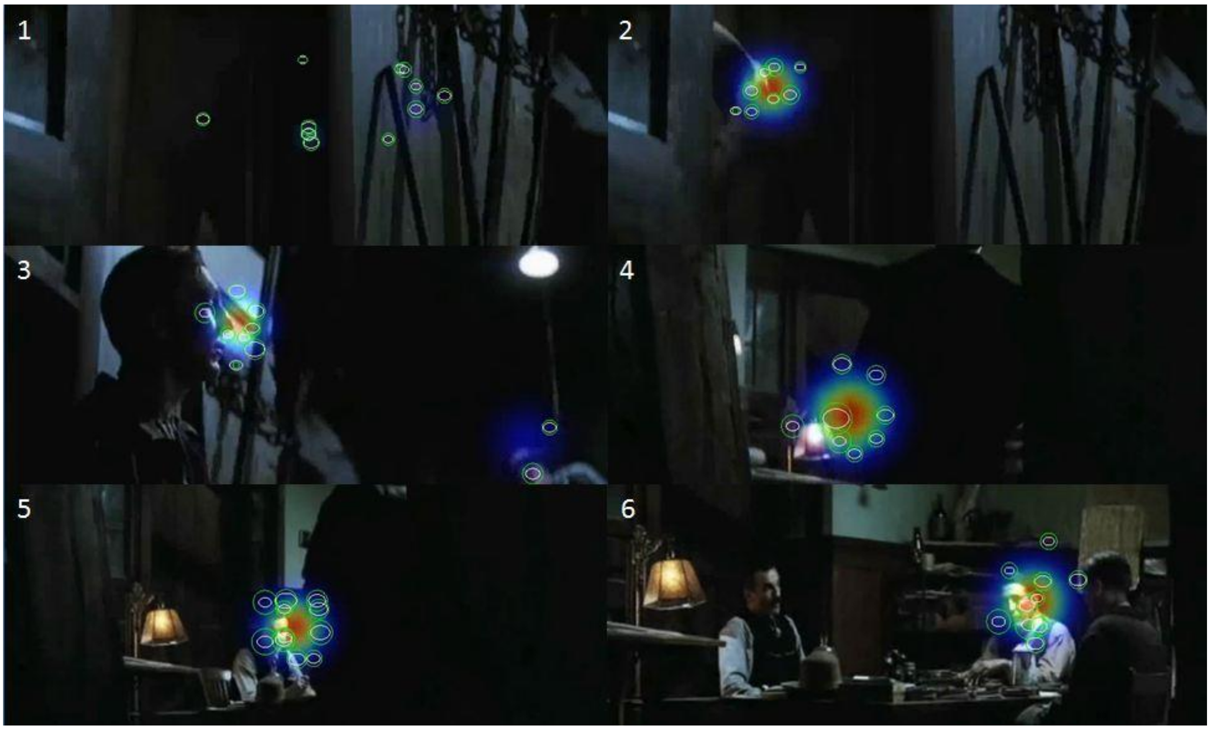

Figure 2

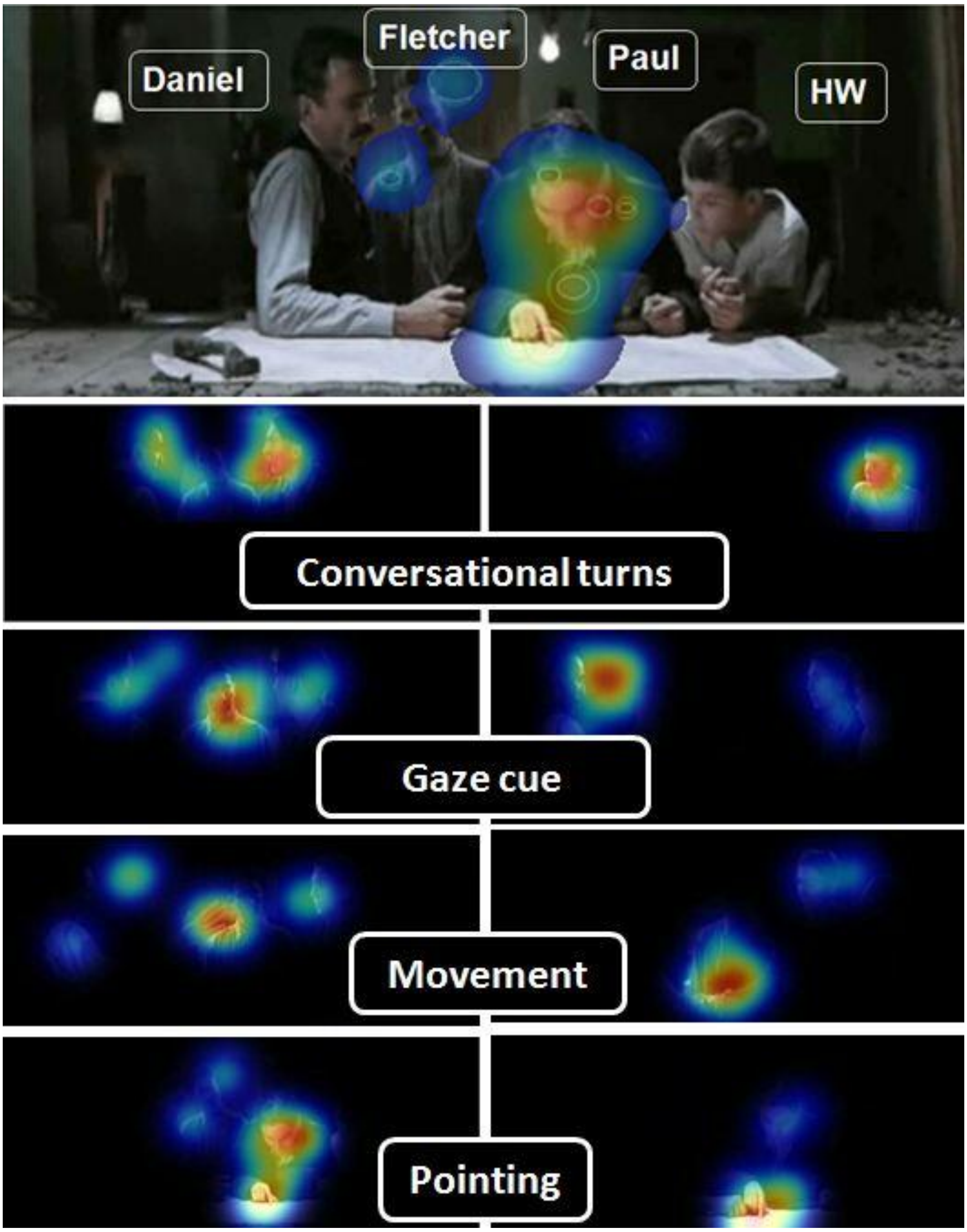
Daniel
Fletcher
Paul
HW
Conversational turns
Gaze cue
Movement
Pointing

Figure 3

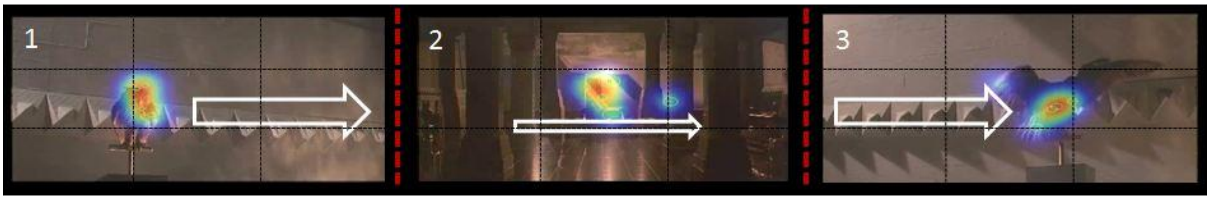

Figure 4

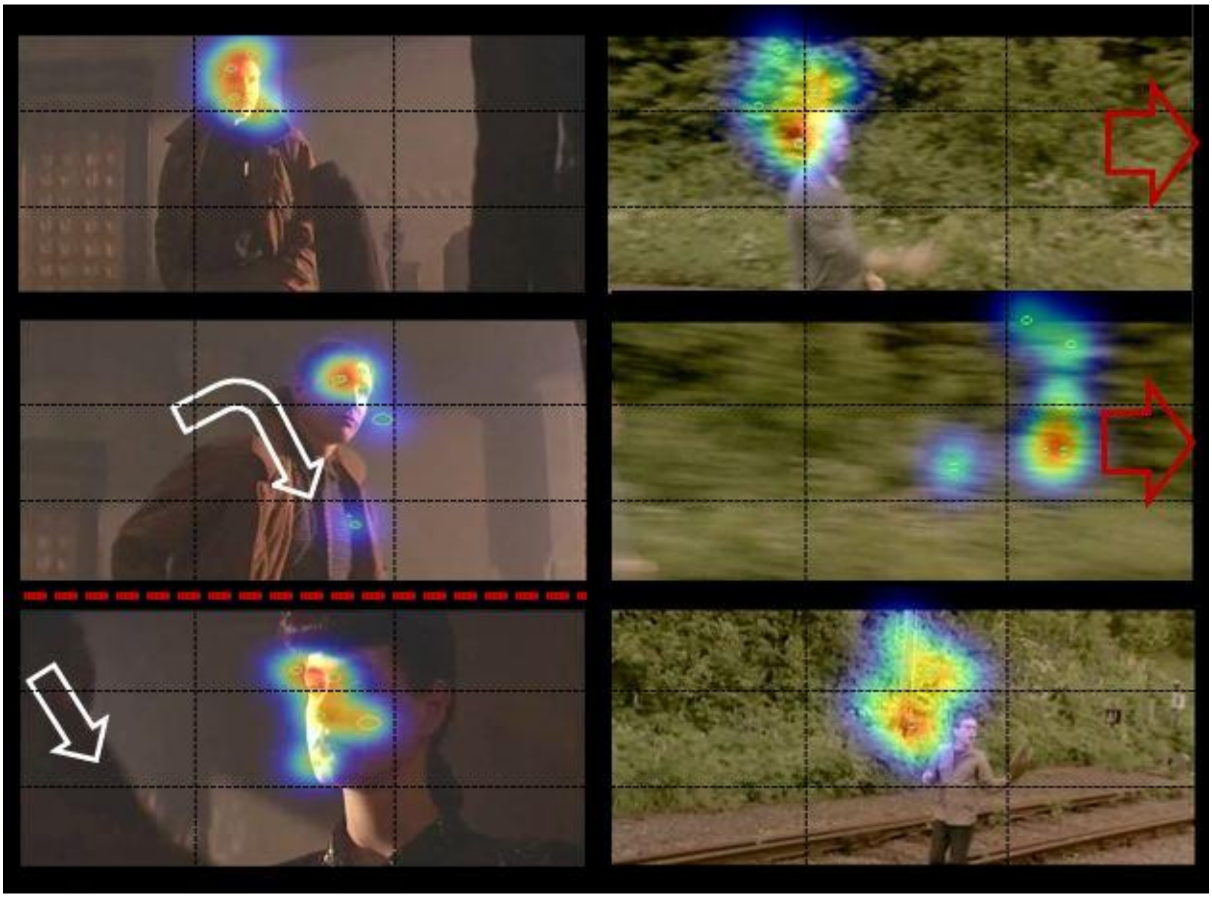

Figure 5

Figure 6

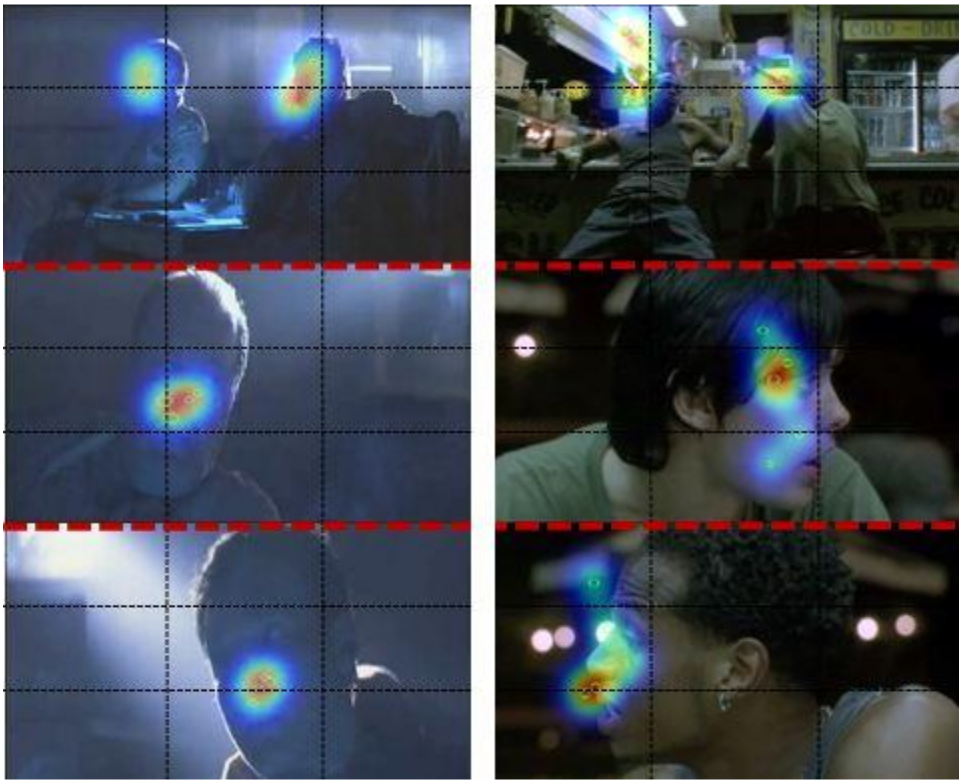

Figure 7

Figure 8

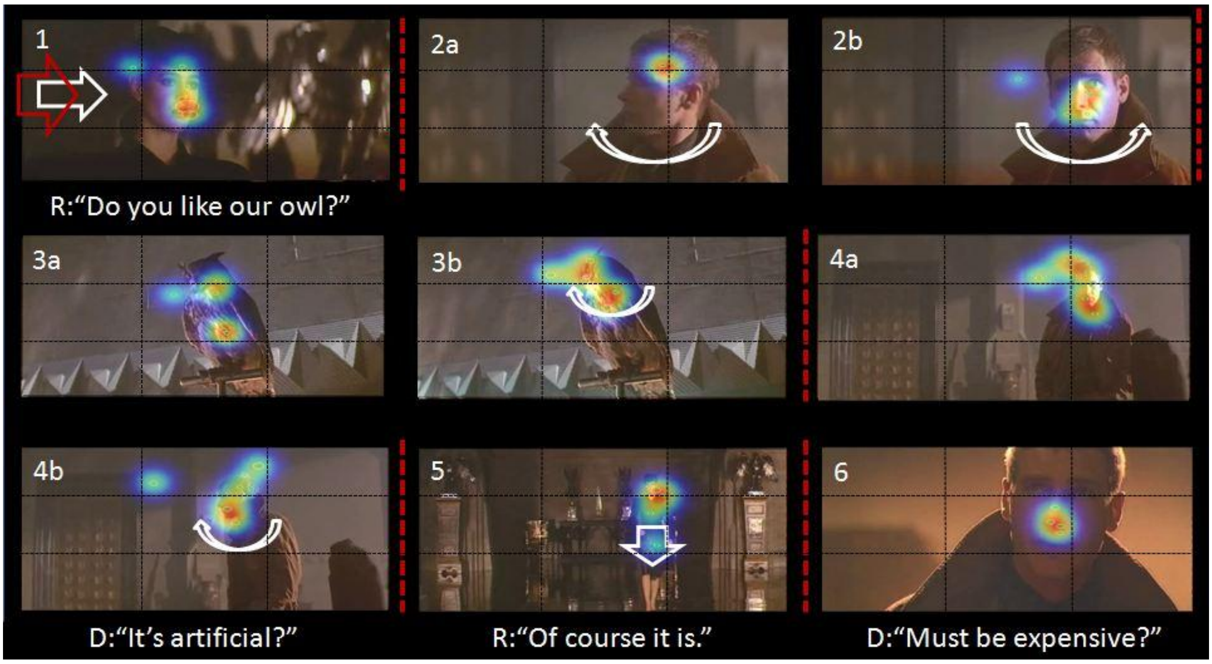
1
2a
2b
R:"Do you like our owl?"
3a
3b
4a
4b
5
6
D:"It's artificial?"
R:"Of course it is."
D:"Must be expensive?"

Figure 9