A Window on Reality: Perceiving Edited Moving Images
Tim J. Smith, Daniel Levin and James E. Cutting
Current Directions in Psychological Science 2012 21: 107
DOI: 10.1177/0963721412437407

The online version of this article can be found at:
http://cdp.sagepub.com/content/21/2/107
We spend a fifth of our waking lives watching movies, television, or some other form of edited moving image (U.S. Bureau of Labor Statistics, 2010). Nonetheless, until recently, research exploring the cognitive and perceptual bases of film has been a lonely activity. Near the inception of film, some astute psychological analyses appeared (Arnheim, 1957; Münsterberg, 1916/1970; Pudovkin, 1929), but interest waned. Over the last 50 years, only a handful of researchers (Carroll & Bever, 1976; Hochberg & Brooks, 1978; Kraft, 1987) have explored the profound questions underlying our near-universal ability to understand film. Part of the reason for this lack of interest has been the presumption that the discontinuous stream of images characteristic of film renders it utterly different from real-world perception and understandable only as a case of media-specific learning (Messaris, 1994). However, recent film theory, a renewed interest in research exploring natural events, and the recent availability of suitable technical and conceptual tools have conspired to end this drought. In this article, we review this research and argue that it supports a cognitive science of film viewing. The research makes clear how film constitutes a form of focused continuity that maintains consistency in cues that natural vision relies upon while allowing discontinuity in features and events that vision does not countenance.

Continuity Editing Reflects Natural Shifts of Attention Within Scenes

In the 1890s, most films depicted simple real-world scenes or staged narratives filmed in a single run (a shot) from a static camera. Shortly thereafter, filmmakers combined multiple shots—separated by cuts (sharp discontinuities from one shot to the next), dissolves, fades, or wipes—to create more compelling visual narratives. Through trial and error, a suite of staging, filming, and editing conventions emerged that allowed viewers to effortlessly integrate these diverse views onto a scene. Most of these conventions, known as continuity-editing rules (or Hollywood style), were in common usage by 1918 (Bordwell, Staiger, & Thompson, 1985; Bordwell & Thompson, 2001). They permeate much of visual media to this day.

Filmmakers believe that by adhering to continuity-editing rules they can make a cut “invisible” (Dmytryk, 1986) and ensure that “the spectator’s illusion of seeing a continuous piece of action is not interrupted” (Reisz & Millar, 1953, p. 216). For example, a typical scene often begins with a shot of the location and the characters within it (an establishing shot; Fig. 1a), then progresses to medium shots framing the characters involved in the action, close-ups of objects of interest (Figs. 1c, 1d, 1e, & 1f), or shots of the speaker over the shoulder of the listener (Fig. 1b). This gradual honing in allows an unfolding of important details in a manner similar to how an observer attends in the real world (Münsterberg, 1916/1970). This observation has been recently confirmed by the application of high-speed infrared eye tracking to dynamic scenes (Mital, Smith, Hill, & Henderson, 2011). Because of acuity limitations, a viewer must move his or her eyes (perform a
saccadic eye movement) to see a part of a visual scene in detail. This action projects the external area of interest onto the most sensitive part of the retina, the fovea. After a saccade, the eyes are relatively stable (fixated), and visual encoding can occur. The sequence of saccades and fixations made during a dynamic scene mirrors the honing in to a scene typically seen within the Hollywood style (Smith, in press). Viewers attend to areas of high motion (Mital et al., 2011), faces of people engaged in conversation, and objects relevant to the viewing task (Smith & Henderson, 2008).

The timing of these shifts is yoked to the timing of cuts, and vice versa. For example, a sudden onset of motion—such as a character’s head turn or hand movement—attracts the attention of the viewer and impedes the ability to detect disruptions (Levin & Varakin, 2004). A film editor often uses the same motion to cut to a different view of the scene (a match-action cut; Fig. 2). This gives the impression of continuous flow and minimizes viewer awareness of the cut. In a cut-detection task, Smith and Henderson (2008) demonstrated that a third of all match-action cuts were missed, whereas only a tenth of cuts between scenes were missed (Fig. 2). Continuity-editing rules also use other natural attention cues, such as conversational turns, character gaze shifts, and pointing gestures (Smith, in press). By piggybacking on natural visual cognition, Hollywood style presents a highly artificial sequence of viewpoints in a way that is easy to comprehend, does not require specific cognitive skills, and may even be understood by viewers who have never watched film before (Schwan & Ildirar, 2010).

**Statistical Regularities in Visual Features Within and Across Scenes**

Contemporary films generally have shorter-duration shots and more motion—they are also darker than earlier films (see Fig. 3; Smith et al., 2012). Between and within scenes, cuts are more likely to occur when the visual field changes in a discontinuous way (Smith et al., 2012; Smith & Henderson, 2008). Match-action cuts, which occur when the viewer perceives the continuation of a visual field, are less likely to be missed than cuts that occur between different scenes (T. J. Smith & J. M. Henderson, 2008, Journal of Eye Movement Research, 2, p. 4). Copyright 2008, Journal of Eye Movement Research. Adapted with permission.
Since about 1960, the shot lengths of films analyzed as sequences across entire films have increasingly approached a \(1/f\) pattern as well (Cutting, DeLong, & Noltehelfer, 2010). Given that films occupy our minds and drive attention, it seems fitting that the shot-duration patterns of popular film might increasingly be like those that our minds may naturally generate.

**The Common Experience of Visual Narrative**

Driving the attention of a single viewer is insufficient; mainstream film must work for all. When multiple viewers attend to the same static scene, they do not necessarily look in the same places at the same time (Mannan, Ruddock, & Wooding, 1997). It might be expected that an increase in complexity of a dynamic scene would also increase differences among viewers. However, the opposite is true. The gaze of viewers is highly coordinated in dynamic scenes, a phenomenon called \textit{attentional synchrony} (see Fig. 4; Smith, 2006). This synchrony is manifest as the spontaneous clustering of the gaze positions across multiple viewers when free-viewing dynamic naturalistic scenes or edited sequences (Mital et al., 2011).

Synchrony of viewer experience was hypothesized by the film editor Walter Murch (2001). While editing \textit{The Conversation} (Francis Ford Coppola, 1974), he noticed that he often felt compelled to blink when Gene Hackman’s character blinked on screen and also to select that moment to cut. He hypothesized that the synchronization of blinks between the entire audience and the actors would entrain viewers’ thoughts to the narrative. Blink synchrony between viewers during film viewing was recently confirmed by Nakano and colleagues (Nakano, Yamamoto, Kitajo, Takahashi, & Kitazaw, 2009). They showed that viewers delayed blinks until the end of events and mirrored blinks of a speaking actor with a delay of...
250 to 500 milliseconds (Nakano & Kitazawa, 2010). As Murch (2001) suggested, blinks may indicate a perceptual break in communication that an editor can use to hide a cut. However, explicit measurement of the co-occurrence of viewer eye blinks with cuts during a cut-detection task indicated that most cuts did not coincide with blinks, although a significant number of within-scene cuts that were missed did (11.6%; Smith & Henderson, 2008). Blink synchronization may be evidence of attentional synchrony across viewers and entrainment to certain features including dialogue and narrative events.

This synchronization of viewer experience has also been found in functional magnetic resonance imaging (fMRI). Hasson, Nir, Levy, Fuhrmann, and Malach (2004) showed the opening 30 minutes of The Good, the Bad and the Ugly (Grimaldi & Leone, 1966) to participants in a scanner. Interobserver comparisons of brain responses during free viewing of the movie revealed that activity in 45% of the neocortex was highly similar across viewers. These areas included regions involved in vision, hearing, language processing, emotion, and multisensory integration. This neural similarity was also shown to increase with the degree of attentional synchrony (Hasson et al., 2008).

A similar synchrony has been observed in how viewers perceive event structure depicted in film and real-world dynamic scenes. If multiple viewers are shown a video of a simple activity, such as folding laundry, and are asked to press a button when one meaningful event ends and another begins, there will be considerable agreement (Newtson, 1973). Event segmentation is believed to be critical for the efficient distribution of cognitive resources, optimizing the organization in space and time of key features in memory (Zacks, 2010).

Spontaneous event segmentation has also been shown to occur while watching edited film sequences (Zacks, Speer, Swallow, & Maley, 2010). Viewers viewed the classic children’s film The Red Balloon (Larmorisse, 1956) while their brain activity was recorded. Afterward viewers segmented the film into events. Cuts that coincided with large changes in action both were identified as event boundaries and induced brain activity in areas of the cortex associated with motion processing (the human Medial Temporal area; MT) and attentional control (dorsal frontal areas). By comparison, cuts that simply changed viewpoint within a scene while maintaining continuity of time, location, and action were not identified as boundaries either behaviorally or by brain activity. Instead, such cuts produced activity in cortical areas associated with attentional control and task switching, possibly confirming the role of attentional shifts in obscuring cuts (Smith, in press).

These studies suggest that the perception of dynamic scenes entails formulating minimal expectations about event sequences and distributing attention appropriately. As long as these expectations are satisfied, continuity will be perceived. This focused continuity prioritizes only the audiovisual features of a scene that are important to the viewer at the moment, allowing irrelevant features such as background details to change without being noticed or violating a priori assumptions of continuity (Smith, in press). During film viewing, these expectations are preserved by editing conventions, allowing the viewer to orient quickly to new shots and to check whether their expectations are satisfied. Future investigations of event
perception during film viewing should focus on uncovering the extent and form of these expectations. These may in turn reveal how malleable such expectations during real-world event perception are.

The Representational Basis of Film

Editing an action at times when attentional shifts would naturally occur helps ensure that film viewing is a lot like viewing the real world. But how is it possible that the sudden change in viewpoint does not lead to confusion? A cut from a medium shot to a close-up (Figs. 1b &1c) creates a sudden change in viewpoint, enlargement and rotation of characters, and omission of many scene details. Such a transition in a real-world scene would entail eye, head, and body movement over an extended period of time, all of which provide feedback on the extent of the movement enabling the viewer to update his or her spatial representations of the scene (Tatler & Land, 2011).

One answer to this question arises naturally from the focused-continuity hypothesis, which not only emphasizes how film is in some ways perceptually consistent but also makes clear that other forms of consistency are not particularly important. Thus, we may lack expectations about object location and appearance that might be violated by the change in viewpoint. Circumstantial evidence in support of this hypothesis comes from continuity errors. Often, striking changes in the appearance and placement of objects across cuts go unnoticed by viewers. Levin and Simons (1997) found that most viewers failed to report a change in identity of an actor across a match-action cut or a change in costume or props across a shot/reverse-shot sequence during a dialog (similar to the change in scarf color between shots a and b in Fig. 1). However, change blindness was not absolute; thirty-three percent of viewers spontaneously reported seeing such changes, and the rate of detection increased when the object was of greater interest. These results suggest that we do not mentally represent all details of a scene; instead, we represent only enough to follow the narrative.

But what form do these representations take? If we were to form a coherent three-dimensional, allocentric (map-like) representation of the depicted scene, it would entail mentally rotating each camera viewpoint and matching details across shots. This would seem to require significant mental effort. Recent empirical investigations of space and object memory across cuts suggest that we represent edited sequences differently than we represent the real world. During real-world tasks involving navigation of the environment, we often shift our attention to objects that were previously fixated but are currently beyond the visual field, such as saccading to a kettle that is behind us when making tea (Tatler & Land, 2011). Thus, we appear to retain at least rudimentary allocentric representations of task-relevant object locations. These locations allow us to maintain temporarily limited information concerning appearance and identity.

Across a cut, memory for object location in a scene is significantly worse than memory for object identity or color (Hirose, Kennedy, & Tatler, 2010). Memory for relations among locations is also poor even after many hours of exposure to the same set (Levin, 2010). Only when a scene is repeatedly presented from the same viewpoint, as occurs for sitcoms filmed in front of studio audiences, are viewers able to accurately remember the spatial relationship between locations (Levin, 2010; Fig. 5). This evidence suggests that an allocentric representation of scenes is not generally retained during film viewing. Future research can determine whether this is due to a failure to represent spatial information or to the use of a different reference frame (e.g. egocentric, viewpoint-dependent) and how featural information within it is associated with these unreliable locations.
Summary
In this article, we have reviewed empirical investigations in film perception and the cognitive foundations of the dominant conventions of film form known as the continuity-editing rules. Film has existed for over a century, and early psychologists wondered how we perceive edited moving images (Münsterberg, 1916/1970). However, it is only with the recent advent of new methodologies in cognitive psychology—such as eye tracking and fMRI—that we have a new willingness to deal with more naturalistic stimuli in the study of visual cognition. Thus, researchers are beginning to understand the cognitive processes involved in that universal pastime: watching movies.

Recommended Reading


Smith, T. J. (in press). (See References). A summary of how we attend to dynamic scenes and a cognitive theory of continuity editing.


Acknowledgments
We would like to thank Jeff Zacks and two anonymous reviews for their constructive comments on an earlier draft of this manuscript. Elements of this review were presented in the “Psychocinematics” symposium organized by Art Shimamura at the 52nd annual meeting of the Psychonomic Society, November 2011, and at past meetings of the Society for Cognitive Studies of the Moving Image (http://scsmi-online.org/).

Declaration of Conflicting Interests
The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding
Tim J. Smith was partly funded by Leverhulme Trust (Ref F/00-158/BZ) on the Dynamic Images and Eye Movement project (http://thediem project.wordpress.com/).

References


