

# Memory-guided attention during active viewing of edited dynamic scenes

**Christian Valuch**

Faculty of Psychology, University of Vienna, Austria  
Georg-Elias-Müller-Institute for Psychology,  
Georg-August-University Göttingen, Germany



**Peter König**

Institute of Cognitive Science,  
University of Osnabrück, Germany  
Department of Neurophysiology and Pathophysiology,  
University Medical Center Hamburg-Eppendorf, Germany



**Ulrich Ansorge**

Faculty of Psychology, University of Vienna, Austria



Films, TV shows, and other edited dynamic scenes contain many cuts, which are abrupt transitions from one video shot to the next. Cuts occur within or between scenes, and often join together visually and semantically related shots. Here, we tested to which degree memory for the visual features of the precut shot facilitates shifting attention to the postcut shot. We manipulated visual similarity across cuts, and measured how this affected covert attention (Experiment 1) and overt attention (Experiments 2 and 3). In Experiments 1 and 2, participants actively viewed a target movie that randomly switched locations with a second, distractor movie at the time of the cuts. In Experiments 1 and 2, participants were able to deploy attention more rapidly and accurately to the target movie's continuation when visual similarity was high than when it was low. Experiment 3 tested whether this could be explained by stimulus-driven (bottom-up) priming by feature similarity, using one clip at screen center that was followed by two alternative continuations to the left and right. Here, even the highest similarity across cuts did not capture attention. We conclude that following cuts of high visual similarity, memory-guided attention facilitates the deployment of attention, but this effect is (top-down) dependent on the viewer's active matching of scene content across cuts.

## Introduction

Feature films, TV shows, newscasts, video clips, and most content from video-sharing websites are instances of edited visual dynamic scenes. These are consumed on

a daily basis by many individuals. The relevance and pervasiveness of edited material has been growing in recent years, since Internet streaming on laptops, smart phones, and other mobile devices became broadly available. In spite of these developments, relatively little experimental work has been devoted to understanding how the human visual system processes such edited material. A general characteristic of edited material is that it contains cuts every couple of seconds, which are abrupt visual changes from one shot to another. Such cuts may appear within a scene or jointly with a change of scene. Since the early days of cinema, the average shot length between two cuts declined from more than 12 s to below 4 s in contemporary movies (Cutting, DeLong, & Nothelfer, 2010). Cuts serve as a ubiquitous artistic tool that film and media editors use to effectively transport the narrative (e.g., Bordwell & Thompson, 2001; Frith & Robson, 1975; Hochberg & Brooks, 1996; Murch, 2001). Despite cuts, viewers often perceive related shots as part of the same narrative element (Cutting, Brunick, & Candan, 2012; Zacks, Speer, Swallow, & Maley, 2010). Given the pervasiveness of edited visual material, it is important to fully understand how edited videos can be perceived effortlessly, even though cuts confront the viewer with the frequent and relatively drastic visual discontinuities.

In professionally produced media, cuts usually comply with a range of professional production conventions, known as *continuity editing* (Bordwell & Thompson, 2001; Reisz & Millar, 1968; Thompson, 1999). Cinematic continuity characterizes cuts that are easy to follow for viewers and minimize the potential to

Citation: Valuch, C., König, P., & Ansorge, U. (2017). Memory-guided attention during active viewing of edited dynamic scenes. *Journal of Vision*, 17(1):12, 1–31, doi:10.1167/17.1.12.

doi: 10.1167/17.1.12

Received January 15, 2016; published January 11, 2017

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution 4.0 International License.



Downloaded From: <https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/935953/> on 09/14/2018

distract the viewers' attention and cognitive processing from the unfolding movie events. Although continuity editing guidelines were discussed in film textbooks (e.g., Arijon, 1976; Bordwell & Thompson, 2001; Murch, 2001; Reisz & Millar, 1968), there is no generally accepted definition of cinematic continuity. However, authors agree that continuity edits are often driven by the continuation of the depicted action, usually in a spatiotemporally coherent stream of events (for reviews see Magliano & Zacks, 2011; Smith, 2012). An example are *match-action* cuts: The onset of an action (e.g., a person grabbing a chair) is shown in one shot, and the action's continuation (e.g., the person sitting down) is shown from a different view in the following shot. Such cuts are subjectively perceived as particularly smooth (Shimamura, Cohn-Sheehy, & Shimamura, 2014), and they do not disrupt the perceptual experience of the viewers (Smith & Henderson, 2008; Smith & Martin-Portugues Santacreu, in press): Smith and Henderson's (2008) participants failed to report a surprisingly large proportion of match-action cuts (and other within-scene cuts [WSC]) but cuts that juxtaposed two different scenes (i.e., between-scene cuts [BSC]) mostly correctly detected. While such studies validate that continuity edits minimize distractions from the video content, the responsible cognitive mechanisms for the processing of edited material are a matter of debate (e.g., Berliner & Cohen, 2011; Levin & Simons, 2000; Magliano & Zacks, 2011; Smith, Levin, & Cutting, 2012).

Theoretical discussions of the principles underlying the perception of continuity in edited material generally focus on processes related to attention (e.g., Levin & Simons, 2000; Münsterberg, 1916/1996; Smith, 2012), and processes related to memory and comprehension (e.g., Berliner & Cohen, 2011; Magliano & Zacks, 2011). Here, we tested if relatively low-level perceptual-attentional mechanisms contribute to continuity perception, but recent attempts to theoretically formalize how cognitive processes might mediate continuity across cuts more generally stressed that conceptual cognitive processing is critical. Before we turn to the perceptual-attentional mechanisms, we review these conceptual theories.

The *attentional theory of cinematic continuity* (Smith, 2012) assumes that film viewers are active perceivers who constantly search for information that helps relating successive shots to one another and understanding the overall narrative as it unfolds across successive scenes. Smith suggested that continuity editing works because viewers always selectively attend to only a small subset of the audiovisual features present in a movie scene, and filmmakers insert conceptual cues to guide the viewers' attention and expectations. Examples for such cues can be sounds that announce an upcoming event not yet in view (e.g.,

a telephone ringing, which will be answered in the upcoming shot), dialogues between actors (e.g., the question by an actress generates the expectation of the response by another actor), or pointing gestures or a gaze direction of an actor. These cues could then modify the viewers' attentional top-down set ahead of the cut, such that viewers will be able to effortlessly shift their attention to the expected objects in the novel shot. In contrast, when attention is not cued before the cut, to establish narrative continuity, a viewer would need to engage in a more reflective process and try to identify conceptual characteristics that help retrospectively inferring the relation to the precut scene (Smith, 2012).

Related to the forgoing, *event segmentation theory* (Magliano & Zacks, 2011; Zacks et al., 2010; Zacks & Swallow, 2007) suggests that the perception of continuity depends on how humans segment their visual experience into conceptual units of events. More specifically, the viewer's expectations about how depicted events will unfold could be a crucial determinant of establishing continuity. The general prediction is that viewers will experience continuity across a cut if they instantaneously recognize the postcut scene as a continuation of the event in the precut scene. For example, if one shot shows an actor approaching a car, and the next shot directly cuts to the person sitting in the car and starting the engine, not all subevents (e.g., of actually getting into the car) would need to be depicted for an immediate understanding of the narrative continuity across the cut. Indeed, experiments showed that participants subjectively perceived that one event had ended and a new event had begun if cuts go along with situational and action discontinuities, such as cuts between different scenes (e.g., Cutting et al., 2012; Magliano & Zacks, 2011; Zacks et al., 2010). However, while such cuts often temporally coincide with perceived event boundaries, a cut, by itself, is neither sufficient nor necessary for perceiving an event boundary, because segmenting depends more on the narrative breakpoint—such as the clear end of one depicted activity and start of a new activity—than on the visual discontinuity of the cut (Schwan, Garsoffky, & Hesse, 2000).

Both of the above reviewed theories further suggest that (a limited capacity) working memory is important for establishing continuity across cuts. It could maintain coarse attentional settings and expectations about the upcoming shot (Magliano & Zacks, 2011; Smith, 2012; Zacks et al., 2010). While this view clearly emphasizes the role of conceptual representations for understanding the content of an edited video, it does not preclude a role of visual features for continuity perception. In fact, research showed that stronger visual changes (which usually occur at cuts between two different scenes) can trigger perception of a new

event (Cutting et al., 2012). More generally, visual features are essential for guiding attention (Wolfe, 1994; Wolfe & Horowitz, 2004), and visual working memory content can bias the viewers' attention toward objects that match these memory representations (Hutchinson & Turk-Browne, 2012; Olivers, Meijer, & Theeuwes, 2006). Accordingly, the perception of conceptual relations across cuts could benefit from visual working memory, because attentional selection precedes the recognition of conceptual relations. In accordance with *attentional theory of cinematic continuity* (cf. Smith, 2012), after each cut, attention could be drawn toward scene content matching the viewer's visual expectations, which will often establish a direct relation between two successive shots. Conversely, without certain visual relations between successive shots, viewers will have more difficulties recognizing the conceptual relations between them. In such cases, where the features change more dramatically across the cut, *event segmentation theory* predicts that viewers will perceive the onset of a new scene or event (e.g., Cutting et al., 2012; Magliano & Zacks, 2011; Zacks et al., 2010).

Motivated by these considerations, we set out to test if attentional selection during the very first moments following a cut is influenced by visual feature similarity relative to the preceding shot. Hitherto, it has never been tested if visual feature similarity across the cut (a) facilitates the viewers' covert shifting of attention during film viewing, and (b) is on average higher for one type of continuous cuts (WSC) as compared to one type of discontinuous cuts (BSC). It has also not been tested, (c) how the attentional effect of visual similarity unfolds across time in the postcut images. All three of these novel questions will be addressed in the present study.

Traditionally, in selective visual attention tasks, participants search for a relevant visual target that occurs unpredictably at one of several locations (Posner, 1980). The classic finding is that participants' attention can be cued to a certain location, resulting in shorter reaction times (RTs) and fewer errors if a cue is a valid indicator of the upcoming target's location (Posner, 1980) and its visual properties (Folk, Remington, & Johnston, 1992). Based on these methods, we recently developed an experimental setup in which a relevant target movie is presented next to an irrelevant distractor movie, and, with every cut, the positions of the two movies can switch or stay the same (Valuch & Ansorge, 2015; Valuch, Ansorge, Buchinger, Patrone, & Scherzer, 2014). The participants attentively viewed the target movie and avoided looking at the distractor movie. Whenever the movies switched locations, participants needed to make a saccadic eye movement to the new location of the target movie. Results suggested that the viewers' attention was cued by WSC

(Valuch et al., 2014): After WSC (where the same agent and scene was shown from alternative views before and after the cut), participants relocated their gaze to the new location of the target movie faster than after BSC (where the agent and scene changed across the cut). We hypothesized that this faster attentional orienting could be mediated by a higher similarity of color features across WSC (relative to BSC), because of two established findings from the literature. First, color is known to be a particularly useful feature dimension for guiding selective attention (e.g., D'Zmura, 1991; Wolfe, 1994; Wolfe & Horowitz, 2004). Second, color is a very helpful feature for object and scene recognition in general (Gegenfurtner & Rieger, 2000; Swain & Ballard, 1991). We adapted our target/distractor approach to test if the attentional benefits of WSC critically depend on color features: With every cut, the movies could either continue in color or in black and white ([BW] in a pseudorandomized sequence). To our surprise, the presence of color features indeed facilitated attention shifts after WSC, but this was true even if color was absent in the preceding shot. Accordingly, the color benefit may have had less to do with a memory advantage, and was of a more sensory nature, by enhancing the distinctiveness of the target relative to the distractor movie (cf. Gegenfurtner & Rieger, 2000). At the same time, the benefits of WSC were not fully eliminated despite the absence of color in the postcut shots (Valuch & Ansorge, 2015). Hence, participants might have disregarded color as a memory-relevant feature, because they could never be sure whether the target movie would continue in BW or in color, and instead focus their attention on other features that would allow them to match scene content across cuts more reliably. This possibility also suggests that participants could exert considerable voluntary (top-down) control on their attention. To test these novel hypotheses, we conducted the present study.

The aim of the present study was to thoroughly investigate the contribution of short-term memory for perceived visual features to the attentional benefits of continuity editing. Our hypothesis was that a higher feature similarity across a cut facilitates the viewers' ability to refocus attention on the target movie's continuation, and reduces distractibility by irrelevant visual content. To that end, we tested if visual similarity influences attention above and beyond a pure semantic similarity across a cut. To minimize potentially confounded nonvisual conceptual cues that could contribute to continuity effects (e.g., off-screen sounds, or narrative-driven expectations; cf. Smith, 2012), we used a set of simple action sports videos as stimuli. The target and distractor movies were presented in pairs and easily discriminable by their visual backgrounds and their major visual and semantic contents (e.g., a skiing movie was paired with a surfing movie). Within

each individual movie, the semantic relatedness remained constant across all types of cuts (e.g., a movie about car racing). However, the visual relatedness was higher in WSC (e.g., showing the same car before and after the cut) than in BSC (e.g., showing a different car before and after the cut). We also included a baseline (BL) condition where visual similarity across the cut was close to perfect. Moreover, we computed visual feature similarity to test if it was indeed objectively higher in WSC than BSC. Throughout the full movies, we presented movie pairs either in BW or in color. Different to our previous study (Valuch & Ansorge, 2015) participants always knew whether a movie would continue in color or in BW and could thus safely rely on color features. This allowed us to test if any similarity benefits are further enhanced by the presence of color features. Finally, the present study also investigated if the attentional benefits by visual similarity critically depend on the viewer's goal to attentively follow an extended movie sequence (Experiments 1 and 2), or if they can be found even in brief two-shot sequences in more passive viewing settings (Experiment 3).

## Experiment 1: Do visual similarities influence covert attention after cuts?

Visual attention can be shifted covertly (without eye movements), and if it is shifted, attended information is processed with a higher quality or precision than information at unattended locations (Helmholtz, 1867; Posner, 1980). Covert attention shifts usually precede overt orienting in the form of saccades toward the attended locations (Deubel & Schneider, 1996; Kowler, Anderson, Doshier, & Blaser, 1995). Following cuts, attention shifts might often be covert, which is why mere gaze direction might not always be a reliable indicator of covert attention in movies (cf. Smith, 2012). Here, we tested whether covert attention after cuts is facilitated for postcut information that is not only semantically but also visually similar to the precut shot.

Participants had to attentively follow the target movie and ignore the distractor movie. At each cut, both movies were briefly interrupted, and, after a short blank interval, continued with the first frames of the next shot at either the same locations or switched locations. To measure how visual similarities influenced covert attention, simultaneously with the onsets of the postcut shots, we flashed one digit on the target movie and a different digit on the distractor movie. After every cut, participants reported the identity of the digit

on the target movie. Due to the short presentation of the digits, viewers needed to keep their gaze at screen center to efficiently monitor both potential target locations following a cut. We predicted that the stronger visual similarities of WSC (and BL cuts) would allow participants to deploy their attention more effortlessly to the target movie's continuation. With higher degrees of visual similarity, we expected an increased rate of correct responses (hit rate) and a shortened manual RT (i.e., the time between the onset of the movies and the manual responses). In addition, we also analyzed how the repetition of the target movie's location (as compared to its switch), and the presence of color in the movies modulated viewers' performance.

## Method

### Participants

Twenty-four (20 female) undergraduate psychology students with a mean age of 21 years (range 18–28 years) participated in exchange for course credit. All had normal or corrected-to-normal visual acuity and intact color vision. This experiment and the following experiments were conducted in accordance with the Declaration of Helsinki and American Psychological Association ethical standards. Before the experiments, participants read and signed an informed consent form.

### Movie stimuli

The current study used the same set of movie stimuli as Valuch and Ansorge (2015). Source footage from DVDs available in the public library of Vienna and in several online video databases was prescreened for its usability. The criteria for selection were that footage contained WSC and BSC. It was easy to recognize two subsequent shots as belonging to the same sports movie as they showed the same type of sports. Altogether, we selected 20 different sports movies. In a second step, the movies were sorted into pairs of clearly distinguishable sports. These were: (a) skiing versus surfing; (b) rally racing versus roller skating; (c) BASE jumping versus kayaking; (d) BMX racing versus skateboarding; (e) freestyle skiing versus windsurfing; (f) skydiving versus downhill mountain biking; (g) Nordic walking versus climbing; (h) snowmobiling versus snowboarding; (i) Formula 1 versus bike racing; and (j) figure skating versus Parkour.

To synchronize the cuts, each movie pair was re-edited using Premiere Pro CS5 (Adobe Inc., San José, CA), such that, for example, whenever there was a cut in the skiing movie, there was also one in the surfing movie. This was achieved by searching the source footage for suitable movie segments, keeping original

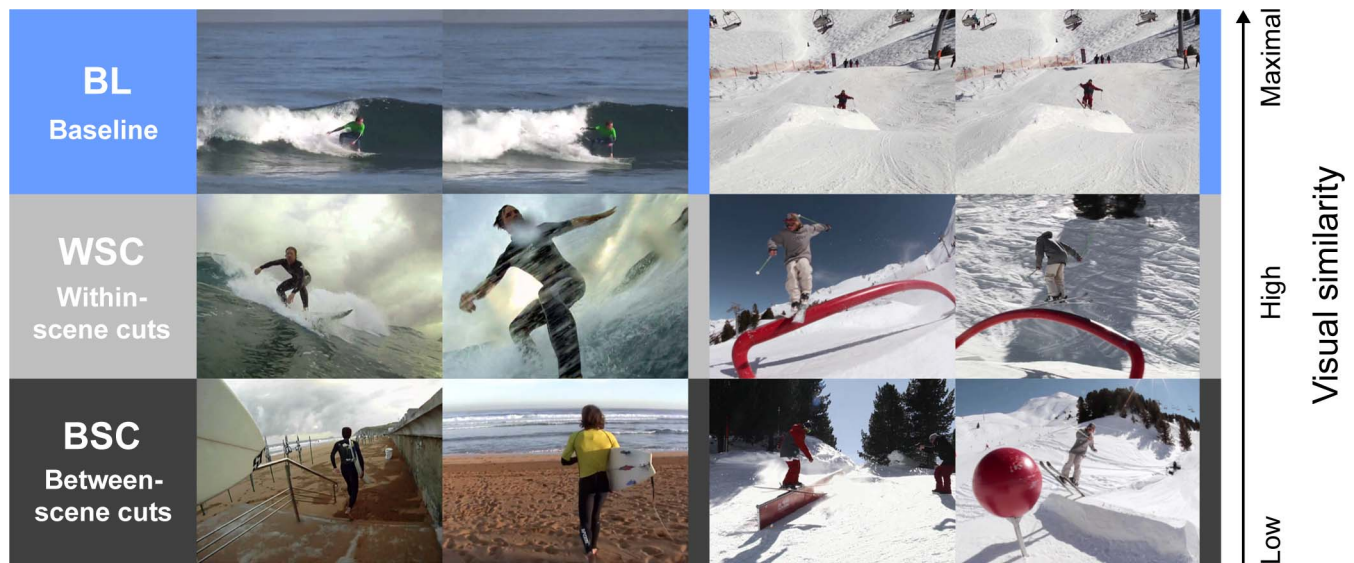


Figure 1. Cut type conditions in Experiments 1, 2, and 3. We manipulated visual similarity across the cut by using BSC, with low visual similarity, or WSC, with a higher visual similarity. BSC presented the viewer with different situations before and after the cut, whereas WSC showed the same actors and situations before and after the cut. As a BL condition with maximal visual similarity, we included mere interruptions of an ongoing scene. Illustration based on still frames of source videos licensed under CC BY 3.0 from Nicolas Fojutu and QParks. For examples of the stimulus material also see Valuch and Ansorge (2015).

WSC and introducing novel BSC, mostly by rearranging the movie's editing sequence. If necessary to synchronize the videos of one pair, pre- and the postcut scenes were trimmed (by deleting video frames adjacent to the cut). Due to the variety of original footage, numbers of suitable BSC and WSC varied across the 10 different movie sets, but across the set of stimuli, we implemented 152 different cuts of each type. In the experiment, each movie (and cut) was shown two times: one time as a target, and a second time as a distractor movie. The order in which a particular movie served as target versus distractor was counterbalanced across participants (see also Procedure and design).

As all movies were paused and screens blanked, we were able to include a BL, in which movies were simply interrupted during an ongoing shot and the same shot continued with the next frame after its interruption (see Figure 1). The movie segments between cuts were of variable lengths, with a mean of 4 s ( $SD = 1.6$  s). All movies were encoded at a resolution of  $400 \times 300$  pixels at 25 progressive frames per second using the H.264 MPEG-4 codec. The movies' apparent size in the experiment was  $11.4^\circ \times 8.6^\circ$  and their on-screen locations were horizontally offset by a distance of  $8.5^\circ$  between movie and screen center. The screen background was black, so that only the movies and a small white fixation circle at screen center were visible.

### Apparatus

Stimuli were presented on an Acer B193 19-in. TFT monitor connected via DVI to a GeForce GT220

(NVIDIA Corp., Santa Clara, CA) graphics card. Screen resolution was set to  $1280 \times 1024$  pixels, with a vertical refresh rate of 60 Hz. The viewable screen area was  $30 \times 37.5$  cm and viewing distance was fixed at 57 cm using chin and forehead rests, resulting in an apparent screen size of  $29.5^\circ \times 36.5^\circ$ . Participants' responses were registered as button presses on a standard USB keyboard. The experimental procedure was implemented in MATLAB (MathWorks, Natick, MA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and run under Microsoft Windows XP SP3 on a Dell OptiPlex 980 PC with an Intel Core 2 Duo E7500 CPU and 2GB of RAM.

### Procedure and design

At the start of each experimental session, participants read the instructions on the computer screen. They were told that the experiment consisted of 20 blocks, each with two movies presented side by side. They were carefully instructed to always fixate on a white circle at screen center and only covertly attend to one of the movies (without directly looking at it). The starting location of the target movie was indicated by a green placeholder at the beginning of each block (see Figure 2). Participants were also told to ignore the other (distractor) movie. The distractor movie's starting location was indicated by a red square at the beginning of the block. Participants were told that the movies would often stop for a short time before they would resume at the same or at switched locations. Participants were informed that movies would switch

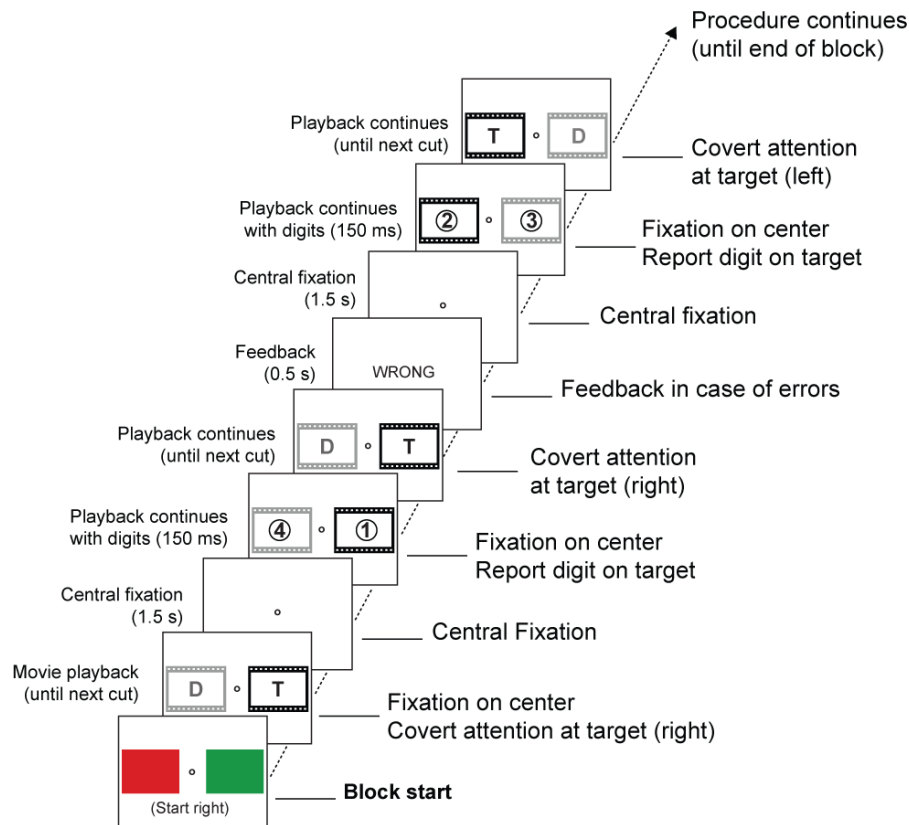


Figure 2. Schematic experimental procedure in Experiment 1. Participants attended to a target movie (T) throughout a whole block in which the target movie always showed the same type of sports, while the distractor movie (D) always showed a different type of sports. The target started at the location indicated by a green rectangle (here: right). Participants ignored the distractor movie that started at the location indicated by a red rectangle (here: left). Movies were interrupted and continued at same or at switched locations. Participants tracked the target movie to manually report its superimposed digit following the interruptions. Dependent variables were hit rate (percentage of correct responses) and manual RT.

locations frequently after these interruptions (overall, location switches occurred in half of the trials).

The participants were further instructed that every time the playback of the movies resumed, there would be two digits briefly flashed on top of the target and distractor movies. The task of the participants was to manually report the identity of the digit presented on top of the target movie. For that, it was necessary to quickly recognize at which location the target movie reappeared because two out of four possible digits that were not identical were shown. Each interruption lasted 1.5 s, immediately followed by the playback at same or at switched locations. For 150 ms following each interruption, the center of each of the two movies was superimposed with a white circular disk of  $2.85^\circ$  diameter and a black digit (1, 2, 3, or 4) in regular Arial font with a height of  $1.5^\circ$  centered on the disk. Participants were asked to quickly and accurately report the digit on the target movie by pressing a preassigned button. At the start of each block, participants placed their left middle and index fingers on the “D” and “F” buttons for responding to 1 and 2,

respectively, and the right index and middle fingers on the “J” and “K” buttons for responding to 3 and 4 on a standard QWERTY keyboard. (The key mapping was explained in the instructions and was the same throughout the experiment.) In case of errors, the word “WRONG” was displayed at screen center for 0.5 s when the movies stopped for the next interruption.

The experiment featured a crossed three-factorial design, with the variables (a) cut type (BSC vs. WSC vs. BL), (b) location (same vs. switched), and (c) color rendition (color vs. BW). Cut types and locations of the movies varied multiple times within a block, whereas color rendition was manipulated block-wise (whole blocks of two movies were either shown fully in color or fully in BW). For the analysis, we considered each cut as one experimental trial, so that the experiment consisted of 456 trials (152 for each of the three cut types, crossed with the two-staged variables location and color rendition). The experiment took about 70 min, including briefing and debriefing, and participants’ short self-paced breaks between the blocks.

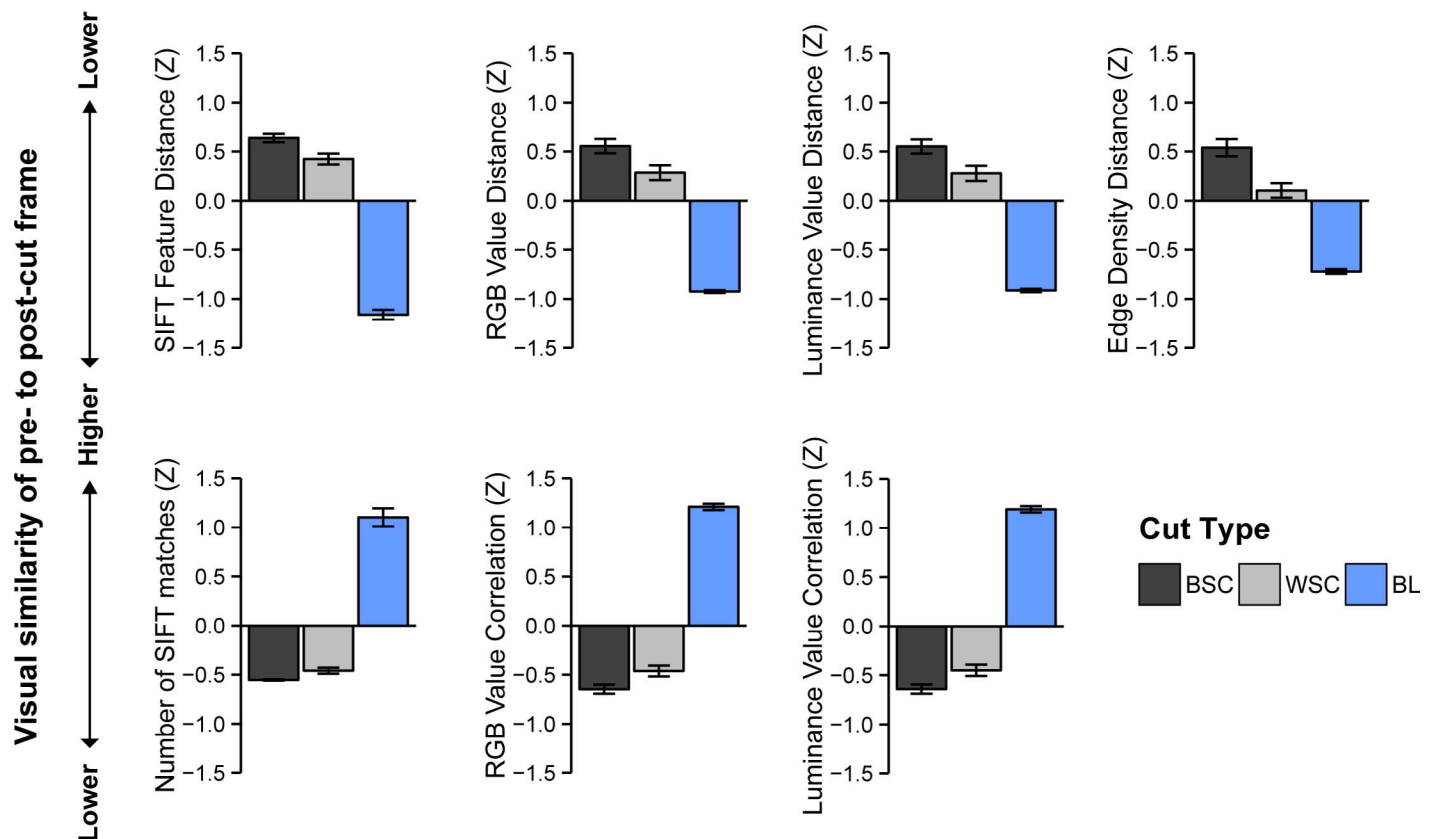


Figure 3. Image similarity across BSC, WSC, and BL. We computed a range of objective similarity metrics to quantify visual similarity between pre- and postcut frames. All metrics were  $z$  standardized for the cuts used in this study. Upper panels show mean-squared Euclidian distances between feature distributions in the two successive frames (smaller values indicate more similarity). Lower panels show mean numbers of matching SIFT keypoints (Lowe, 2004; Vedaldi & Fulkerson, 2010) and mean 2-D-correlations of RGB and luminance values (larger values indicate more similarity). Error bars depict  $\pm 1$  SEM.

### Objective measures of visual similarity across cuts

We compared participants' performance in BSC and WSC against the BL condition. We additionally computationally verified the differences between the cut conditions via objective similarity (or dissimilarity) metrics, and checked if visual similarity was indeed higher in our WSC than BSC. Previous research suggested that metrics corresponding best to human visual similarity judgments and human recognition performance are based on scale invariant feature transform (SIFT) descriptors (Lowe, 2004; Vedaldi & Fulkerson, 2010), as well as on color feature comparisons (Swain & Ballard, 1991). Importantly, such features also explain visual search performance particularly well (Alexander & Zelinsky, 2011). We therefore based our analyses of the visual similarity of pre- to postcut frames on these features. As dissimilarity metrics we computed the mean-squared Euclidian distances between the best-matching SIFT features, as well as red-blue-green (RGB) and luminance feature histograms, and overall Sobel edge densities in the image (see Figure 3, upper panels), whereby smaller values indicate more simi-

ilarity while larger values indicate less similarity. As similarity metrics, we computed the number of matching SIFT keypoints between the pre- and postcut frames (Vedaldi & Fulkerson, 2010), as well as the two-dimensional (2-D) correlations of RGB and luminance values between the pixels in pre- and postcut movie frames (see Figure 3, lower panels). Due to the metrics' different scaling and because we were interested in relative differences between cut type conditions we computed the standardized  $z$  scores for each metric within the whole sample of cuts (see Figure 3). Pairwise comparisons in all metrics verified that similarity was significantly higher in BL than WSC ( $p < 0.05$ ). Also, in all metrics, similarity was significantly higher in WSC than BSC ( $p < 0.05$ ). In addition, Figure 3 shows a large correspondence between the evaluated metrics, indicating that they all validly reflected differences in visual similarity between the cut type conditions.

### Principle component analysis-based similarity index

In addition to our a priori definition of cut type conditions, we wanted to assess if the RT benefits were

Loadings	Variance components						
	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7
<b>Luminance-based PCA</b>							
SIFT matches	0.48	0.29	−0.38	0.38	0.63		
SIFT distance	−0.50	−0.22	0.28	−0.20	0.77		
Luminance distance	−0.42	−0.14	−0.88	−0.19			
Luminance correlation	0.47			−0.87	0.11		
Edge density distance	−0.36	0.92		−0.15			
Standard deviation	1.82	0.82	0.72	0.59	0.40		
Proportion of variance	0.66	0.14	0.10	0.07	0.03		
Cumulative proportion	0.66	0.80	0.90	0.97	1.00		
<b>Luminance- and color-based PCA</b>							
SIFT matches	0.38	−0.19	−0.24	0.59	0.64		
SIFT distance	−0.40	0.17	0.18	−0.44	0.76		
Luminance distance	−0.38	−0.59				0.71	
Luminance correlation	0.41	−0.29	−0.16	−0.48			−0.70
Edge density distance	−0.28	0.24	−0.93				
RGB distance	−0.37	−0.60				−0.70	
RGB correlation	0.41	−0.29	−0.16	−0.46			0.72
Standard deviation	2.17	0.95	0.82	0.73	0.40	0.18	0.05
Proportion of variance	0.67	0.13	0.10	0.08	0.02	0.01	0.00
Cumulative proportion	0.67	0.80	0.90	0.97	0.99	1.00	1.00

Table 1. Results of principal component analyses (PCA) of the different measures for visual similarity between the last precut and the first postcut frame of each cut. *Note:* SIFT = scale invariant feature transform; Comp. = component.

numerically correlated with strength of visual similarity of each particular cut. To reduce the similarity dimensions (as measured by the range of similarity and dissimilarity metrics, see Figure 3) to a single representative index of visual similarity for each cut, we performed a principle component analysis (PCA) on the results of the visual similarity analysis. Because in our experiments, each cut was used either in a BW or color video, we calculated the PCA once for all seven similarity measures (SIFT distance, SIFT matches, RGB distance, RGB correlation, luminance distance, luminance correlation, and Sobel edge density distance), and once for only five of these measures (leaving out RGB distance and RGB correlation). Table 1 depicts the resulting variance components and loadings in these two complementary analyses. As a general characteristic of the PCA method, the first variance component accounts for most of the true variance within the underlying property of interest (in our case “visual similarity”), whereas the later variance components are more vulnerable to noise, which means that they are more strongly influenced by different degrees of precision with which the individual similarity metrics capture the relative differences in visual similarity between individual cuts. Hence, we used only the first variance component as a PCA-based index of visual similarity, which overall explained 66% for the luminance based PCA and 67% for the luminance- and color-based PCA (see Table 1) across all similarity

metrics. This PCA-based similarity index was used to analyze how quickly participants responded depending on the degree of the visual similarity of the cuts in all experiments of the present study. In addition, we verified that this measure validly represented all of the underlying similarity metrics by computing the Pearson correlation coefficients between the original metrics and the PCA-based similarity index (see Table 2). In all cases, the correlations were highly significant, with  $p < 0.001$ , suggesting that our PCA-based similarity index validly reflected differences in all evaluated feature dimensions. For the same cuts, the similarity index computed from the luminance-based PCA corresponded very highly to the result from the luminance- and color-based PCA,  $r = 0.98$ , suggesting that the method captured the relative differences in visual similarity between individual cuts equally well in both versions.

### Data analysis

We collected 10,944 data points (456 trials for each of the 24 participants). For the analysis of hit rates, we used the complete dataset. For the analysis of RTs, we analyzed only correct responses. Both measures were analyzed using repeated-measures analyses of variance (ANOVAs), and, where applicable, Bonferroni-corrected  $t$  tests for post hoc pairwise comparisons. As effect sizes we report generalized  $\eta^2$  in the ANOVA

Similarity metric	Luminance-based PCA	Luminance- and color-based PCA
SIFT matches	0.87 [0.84; 0.89]	0.82 [0.79; 0.85]
SIFT distance	−0.90 [−0.92; −0.89]	−0.87 [−0.89; −0.84]
Luminance distance	−0.76 [−0.80; −0.72]	−0.82 [−0.84; −0.78]
Luminance correlation	0.85 [0.83; 0.88]	0.88 [0.86; 0.90]
Edge density distance	−0.65 [−0.70; −0.59]	−0.60 [−0.66; −0.54]
RGB distance		−0.81 [−0.84; −0.78]
RGB correlation		0.89 [0.87; 0.91]

Table 2. Pearson correlation coefficients for the relation between the principle component analyses (PCA)-based similarity indices and the original similarity metrics which they are based on, with 95% confidence intervals. *Note:* SIFT = scale invariant feature transform; RGB = red-green-blue.

(Bakeman, 2005), and Cohen's  $d$  for pairwise comparisons (Cohen, 1992). In case of significant departures from sphericity (indicated by Mauchly's test), we report  $p$  values after applying the Greenhouse–Geisser correction. Data were analyzed using R Version 3.1.1 (R Core Team, 2015) and ggplot2 (Wickham, 2009). Error bars in plots depict  $\pm 1$  SEM after correcting for between-participants variance (Loftus & Masson, 1994).

## Results

### Hit rates

The analysis yielded a main effect of cut type,  $F(2, 46) = 9.3$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.014$  (see Figure 4A). Pairwise comparisons showed that hit rates were significantly lower for BSC ( $M = 76.8\%$ ) than WSC ( $M = 79.9\%$ ;  $p = 0.004$ ,  $d = -0.29$ ) cuts and BL trials ( $M = 80.9\%$ ;  $p = 0.001$ ,  $d = -0.37$ ). There was no significant difference between WSC and BL. The analysis also yielded a main effect of location,  $F(1, 23) = 62.6$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.101$ , with significantly lower hit rates following location switches ( $M = 74.3\%$ ,  $SD = 16.6\%$ ) than repetitions ( $M = 84.12\%$ ). In addition, a significant interaction of cut type  $\times$  location,  $F(2, 46) = 3.6$ ,  $p = 0.035$ ,  $\eta_G^2 = 0.005$ , and a follow-up simple effects analysis indicated that cut type had a significant effect when the movies switched locations,  $F(2, 46) = 10.8$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.029$ , but not when locations repeated,  $F(2, 46) = 1.2$ ,  $p = 0.319$ ,  $\eta_G^2 = 0.004$ . For switched locations, post hoc  $t$  tests indicated that hit rates were significantly lower for BSC ( $M = 70.5\%$ ,  $SD = 17.44\%$ ) than WSC ( $M = 75.0\%$ ,  $p = 0.010$ ,  $d = -0.42$ ) or BL trials ( $M = 77.3\%$ ,  $p < 0.001$ ,  $d = -0.65$ ), whereas there was no significant difference between WSC and BL ( $p = 0.286$ ). Whether movies were presented in color or BW did not have a significant effect on hit rates,  $F(1, 23) =$

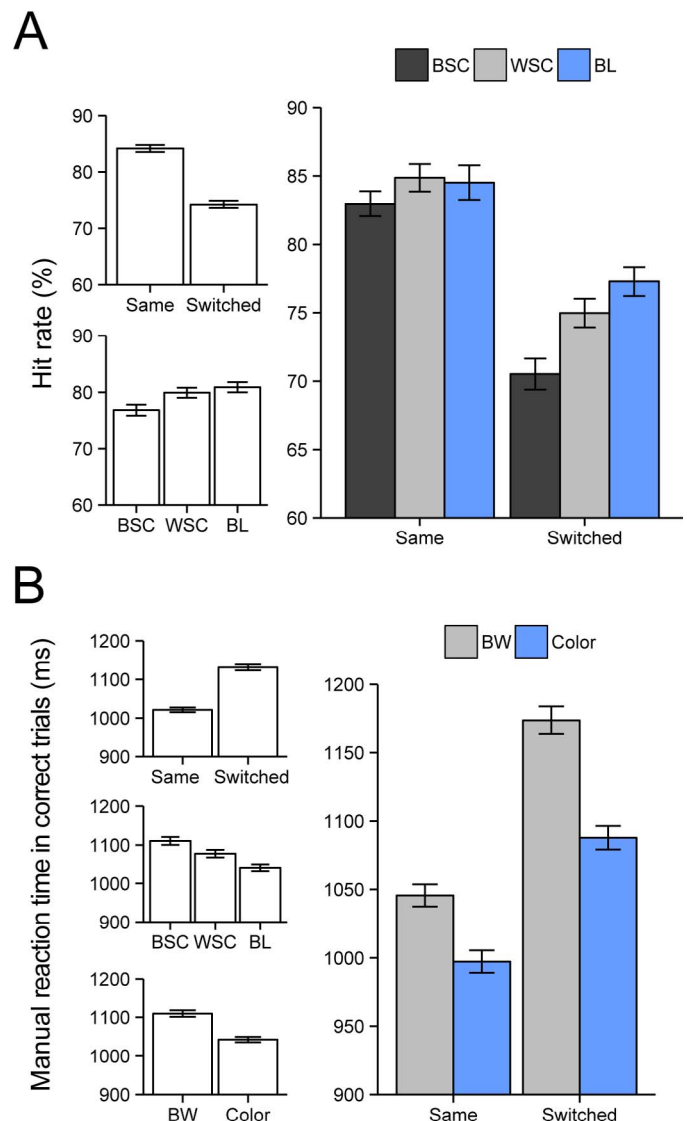


Figure 4. Significant main effects and interactions in the ANOVAs of Experiment 1. (A) Depicted are hit rates (on the ordinate) as a function of location (top left in A: same; switched), cut type (bottom left in A: BSC, WSC, and BL), and of location and cut type (right in A). (B) Depicted are mean correct manual RTs (on the ordinate) as a function of location (top left in B: same; switched), cut type (center left in B: BSC, WSC, and BL), color (bottom left in B: BW; color), and of location and color (right in B). Error bars represent  $\pm 1$  SEM.

0.10,  $p = 0.758$ , and there were no significant interactions between color and any of the other independent variables. We also repeated the analysis with arcsine-transformed hit rates and found a qualitatively identical pattern of significant results.

### Manual RTs

For manual RT analyses, we excluded trials with a wrong or no response (2,270 responses or 20.7%). From the remaining data, we eliminated manual RTs outside

$\pm 2.5$  *SD* of the participant's condition mean RT, keeping a total of 8,481 trials (77.5%). Significant effects are depicted in Figure 4B. First, the analysis of mean correct RTs yielded a significant main effect of cut type,  $F(2, 46) = 27.9$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.024$ . Mean RTs for BSC ( $M = 1111$  ms) were longer than for WSC ( $M = 1078$  ms,  $p = 0.003$ ,  $d = 0.30$ ) or in the BL condition ( $M = 1040$  ms,  $p < 0.001$ ,  $d = 0.67$ ). Moreover, mean manual RTs for BL were shorter than for WSC ( $p < 0.001$ ,  $d = 0.35$ ). Second, there was a significant main effect of location,  $F(1, 23) = 73.9$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.080$ , with shorter manual RTs when the target location repeated ( $M = 1021$  ms) than when it switched ( $M = 1131$  ms,  $d = -1.01$ ). Third, there was a significant main effect of color,  $F(1, 23) = 46.4$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.032$ , with shorter manual RTs for color ( $M = 1043$  ms) than BW ( $M = 1110$  ms,  $d = 0.62$ ) movies. Finally, the analysis yielded a significant interaction of Color  $\times$  Location,  $F(1, 23) = 5.6$ ,  $p = 0.027$ ,  $\eta_G^2 = 0.003$ . Follow-up pairwise comparisons indicated that color was significant, irrespective of whether the locations switched or not (both  $ps < 0.001$ ), but the beneficial effect of color was larger following location switches ( $\Delta RT = 86$  ms,  $d = 0.75$ ) than repetitions ( $\Delta RT = 49$  ms,  $d = 0.46$ ).

### Relationship between visual similarity and RTs

While the analyses of manual RTs in the previous section tested whether the cut type manipulation led to reliable *within*-participant differences in mean RTs (across the number of cuts within each cut type condition), we also wanted to directly relate the actual visual similarity of each cut, as assessed by the objective similarity analysis, with the RTs elicited by this cut (across the number of participants that saw this cut). For the following analysis, we used the median instead of the mean RT because the median is a more robust summary statistic, avoiding that a few participants with particularly fast or slow overall responses exert a strong influence on the result. This is important because this analysis focuses on differences between cuts rather than within-participant RT differences. The visual similarity of a particular cut relative to all other cuts included in the present study was objectively quantified along a range of typical low-level feature similarity dimensions, computed for the last pre-cut and the first post-cut frame (see Figure 3). These different similarity dimensions were reduced to a single similarity index using a PCA approach (see PCA-based similarity index, for details).

In a first step, we fitted a linear model with the PCA-based similarity index and the location factor as fixed effects, separately to the RT data from color trials and BW trials. This was done to verify that visual feature similarity between the pre- and the post-cut shot accounts for part of the variance. In the color trials, the

model resulted in a significant main effect of location,  $F(1, 452) = 44.9$ ,  $p < 0.001$ . In addition, color trials also yielded a significant main effect of similarity,  $F(1, 452) = 21.6$ ,  $p < 0.001$ . The interaction Location  $\times$  Similarity was not significant,  $F(1, 452) = 0.4$ ,  $p = 0.53$ . In line with the previous analyses, the model coefficients reflected increased RTs after location switches,  $b = 116.64$ ,  $SE = 17.40$ ,  $p < 0.001$ , and decreasing RTs with higher degrees of similarity,  $b = -40.59$ ,  $SE = 8.74$ ,  $p < 0.001$ . For the BW trials, the model again yielded a significant main effect of location,  $F(1, 452) = 54.54$ ,  $p < 0.001$ . However, the similarity predictor did not reach significance in this analysis,  $F(1, 452) = 1.97$ ,  $p = 0.16$ , and neither did the interaction of Location  $\times$  Similarity,  $F(1, 452) = 1.62$ ,  $p = 0.20$ . The model coefficients again reflected increased RTs after location switches,  $b = 138.7$ ,  $SE = 18.8$ ,  $p < 0.001$ , and, although not statistically significant, decreasing RTs with higher degrees of similarity,  $b = -19.4$ ,  $SE = 13.8$ ,  $p = 0.16$ .

Next, we plotted the median correct RTs per cut as a function of the degree of visual similarity, separately for the location and the color manipulations. To check if the effect of similarity might have differed between the three cut conditions, we inserted a linear fit into the subdatasets of BSC, WSC, and BL (see Figure 5).

In this plot, two findings are obvious. First, the variability of similarity is not independent of the cut types. A Spearman rank correlation between the similarity indices and the cut type condition (ordered as BSC  $<$  WSC  $<$  BL) suggested a strong correlation between these two variables ( $\rho = 0.76$ ). Second, the slopes of the relationship between similarity and RTs are not homogenous across the three cut conditions. Due to the evident collinearity of similarity and cut type as well as the heterogeneity in similarity slopes for the different levels of cut type, we could not estimate the effects of both these factors within a single statistical model.

To nevertheless assess how important similarity might be for guiding attention in the different cut type conditions, we computed the Pearson correlation coefficients between the PCA-based similarity index and the median correct RT per cut, based on data from all 24 participants (see Table 3). These results illustrate stronger visual similarity was often associated with faster correct RTs in WSC as well as in BSC, although these correlations were not significant for all combinations of the independent variables. In contrast, in BL trials, where all cuts were already highly similar, the residual variance in the similarity measure did not correlate with RTs.

## Discussion

Experiment 1 tested the effect of visual similarity on viewers' covert attention after cuts when viewers are

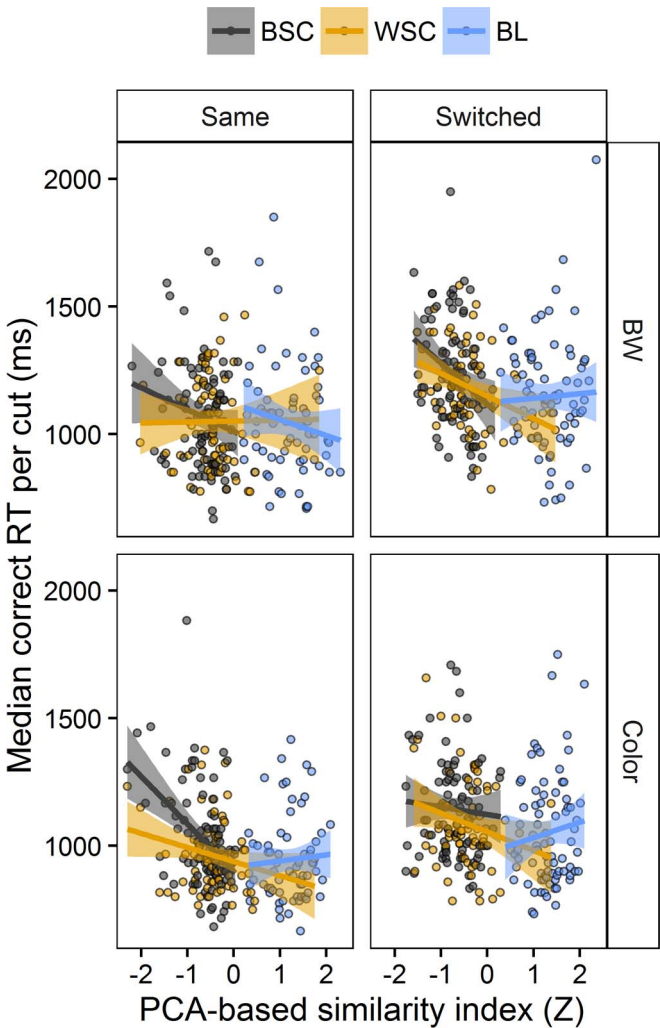


Figure 5. Median correct RT per cut (from 24 subjects) in Experiment 1 as a function of a PCA-based similarity index that represents how visually similar the pre- and the postcut images were. Data are plotted in separate panels for the location and the color factors. Within each panel, colored lines and shaded surfaces ( $\pm 1$  SE) depict a linear fit for subsets of the RT data according to cut type (BSC, WSC, or BL).

actively engaged in following an edited movie on the same topic (here, the same type of sports). Our predictions were that the viewers' hit rates and manual RTs would improve as a function of the degree of visual similarity across the cut (Ansorge, Buchinger, Valuch, Patrone, & Scherzer, 2014). Overall, the results confirmed our predictions and showed gradual benefits in BL and WSC as compared to BSC. In addition to the a priori definition of cuts as BSC or WSC, we also confirmed that WSC were objectively visually more similar than BSC, and that BL cuts achieved the highest similarity scores.

In addition to the beneficial effects of visual similarity, the repetition of the target movie's location (as opposed to a location switch) contributed significantly to participants' performance. Results showed

	BSC	WSC	BL
BW			
Same	−0.18	0.01	−0.15
Switched	−0.29**	−0.30**	0.04
Color			
Same	−0.43***	−0.22*	0.06
Switched	−0.07	−0.23*	0.11

Table 3. Pearson correlation coefficients and  $p$  values for the relationship of similarity and reaction time, as a function of cut type (BSC/WSC/BL), color versus black and white, and location (same/switched). Note: BSC = between-scene cuts; WSC = within-scene cuts; BL = baseline; BW = black and white. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ .

that repeating the target location facilitated responses and hit rates. This effect could have reflected a bias to attend to the previous target location. In the analysis of hit rates, the factor location also interacted significantly with the cut type manipulation: The feature similarity in WSC and BL benefitted mostly trials in which the locations of the movies had switched. Possibly the bias to attend to the previous target location operated independently of a perceptual analysis of the images, so that the influence of cut type was drowned in location-repetition trials. Color effects on RTs were in line with this interpretation. Results showed that color speeded the participant's responses as compared to a BW control condition, and color was particularly helpful following location switches. Importantly, color had an influence that was independent of that of cut type. Maybe color generally increased target-distractor distinctiveness so as to facilitate attending to the target regardless of the particular cut type. This interpretation would be in line with the conclusions of Valuch and Ansorge (2015) where a beneficial influence of color on target movie detection was also found if only the postcut image was colored and a BW image was presented prior to the cut. The current finding of no interaction between color and cut type under blocked color conditions also means that the independence of cut type effects of the presence of color before and after the cut in Valuch and Ansorge (2015) was not due to the fact that, in the previous study, color and BW images were randomized within blocks.

Though Experiment 1 clearly showed that attention was facilitated by visual similarities as the target representations benefitted from WSC, an open question after Experiment 1 is how quickly attention was allocated to target movies and if it was also ever allocated to the distractor movies. One reason for this is that manual RTs only reflected the outcome of several potentially different processing phases preceding the response. As we did not monitor eye movements in Experiment 1, we cannot exclude the possibility that participants—at least in some of the trials—looked at

the target and distractor movie locations. Such eye movements were not instructed and should not have helped to solve the digit discrimination task because the target digit appeared randomly at the left or at the right, simultaneously with target and distractor movie onsets (and the only reliable strategy was to keep fixation at the center and monitor both locations). Yet participants might have made some eye movements toward the movie locations during the period until the next interruption because attending to a movie covertly in the visual periphery for an extended time is not a very natural thing to do. In addition to the possibility of eye movements, the participants' mean RTs were overall relatively long, probably reflecting task difficulty. Importantly, during the time until the response, participants might have quickly shifted their attention to the target movie, or first to the distractor movie and then to the target movie. With the single RT measurement at the end, there is no way to discern between these different possibilities. To address how attention unfolded over time, we therefore conducted a second experiment in which we asked the participants to actually orient their gaze to the designated target movies. Hence, Experiment 2 provided a test of the effect of visual similarities on overt attention after cuts.

## Experiment 2: Do visual similarities influence early eye movements after cuts?

In our second experiment, we asked participants to quickly return their gaze toward the target movie after every cut. In contrast to the manual RTs, the participants' gaze directions were sampled continuously at millisecond resolution, allowing us to look at the time course of attentional deployment toward the target movie. We predicted that cuts with stronger visual similarities would allow participants to deploy their overt attention more effortlessly to the target movie's continuation. Accordingly, cuts with a higher visual similarity should allow for an increased proportion of trials in which the participant's very first saccade landed on the target movie and not on the distractor movie; and a reduced saccadic target RT—that is, the time between the onset of the movies and the initiation of the first saccade to the target movie.

## Method

### Participants

Twenty-four (15 female) new participants with a mean age of 23 years (range 18–37 years) were recruited from the same student population as in Experiment 1.

### Movie stimuli

Because the eye tracker featured a different screen and PC setup, the apparent size of the movies was slightly different from Experiment 1: Each movie appeared at a size of  $9.7^\circ \times 7.1^\circ$  and at a horizontal center-to-center eccentricity of  $7.2^\circ$ . In all other respects, the movies were identical to Experiment 1.

### Apparatus

Eye movements were recorded using an infrared video-based EyeLink 1000 Desktop Mount eye tracker (SR Research Ltd., Kanata, Ontario, Canada) at a sampling rate of 1000 Hz. At the start of each experimental session, a standard 5-point calibration on the observer's dominant eye was performed. At the beginning of each experimental block, a drift check was performed requiring participants to fixate on a centrally presented circle. Drift checks were also conducted every 10th trial, and recalibrations were performed whenever the recorded fixation average was outside a  $1^\circ$  radius of the drift check circle. Movies were displayed on a 19-in. color CRT monitor (Sony Multiscan G400) connected through VGA to a GeForce GT220 graphics card. Screen resolution was set to  $1280 \times 1024$  pixels, with a vertical refresh rate of 60 Hz. The viewable screen area comprised  $35.5 \times 27.4$  cm ( $31^\circ \times 24.2^\circ$ ). The experimental procedure was adapted from Experiment 1, and the EyeLink toolbox (Cornelissen, Peters, & Palmer, 2002) functions were used to trigger the eye movement recording. The display PC had an AMD Athlon Dual Core CPU, 2GB of RAM, running Windows XP SP3. Manual responses were registered via a standard PC keyboard.

### Procedure and design

The design was identical to Experiment 1. Because Experiment 2 was designed to measure overt attention via saccades and fixations, some details of the procedure and the viewing task were adjusted. Importantly, the participants' main task was to look at the target movie as quickly as possible, whenever the playback resumed (after an interruption), and keep it fixated until the next interruption. Because eye movement parameters served as the dependent variables, the digits and their reports were omitted. The procedure within one block is depicted in Figure 6. As soon as the movies started, participants made a saccade to the target movie and kept fixating on it until the next interruption. During the interruption, participants fixated the screen center.

To keep participants engaged, they had to report whether the movies had or had not switched locations after the last seen interruption. To ensure that at each interruption participants returned their fixation to

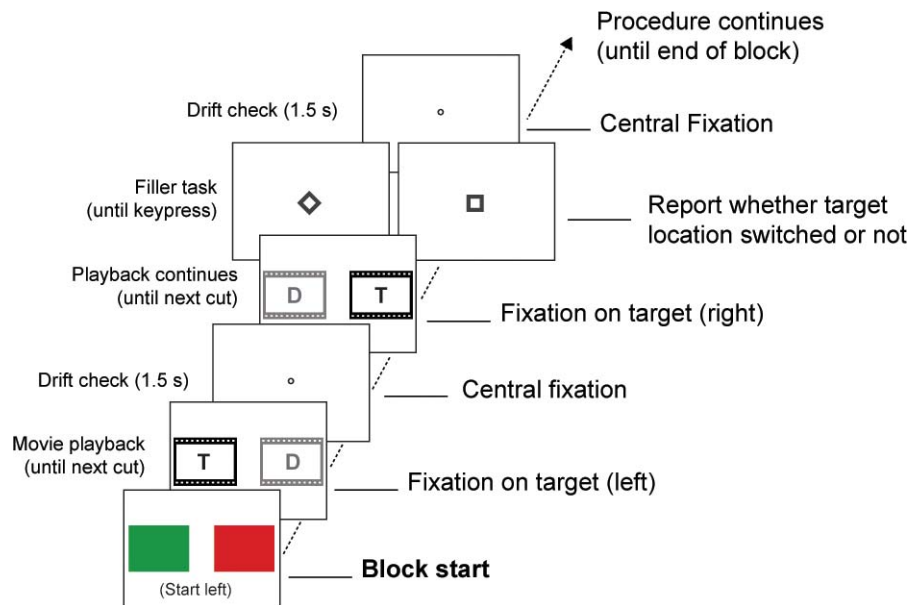


Figure 6. Schematic experimental procedure in Experiment 2. In this example, the target movie started left. After each interruption, participants made saccades toward the target movie and kept fixating it until the next interruption. During interruptions, a filler task was presented to ensure that participants returned fixation to screen center. Dependent variables were the proportion of first saccades going directly toward the target movie, saccadic RT, and fixation durations on the target and the distractor movie, respectively.

screen center, we used either a gray diamond at screen center. For half the participants, squares meant that key #8 had to be pressed for repeated movie locations and #2 for switched locations, whereas this mapping was reversed for the other half of the participants. Participants immediately received negative feedback (“WRONG” displayed at screen center for 0.5 s) if they made an incorrect response. After the button press, the movies’ playback resumed, and participants had to make a saccade toward the target movie and fixate on it as quickly as possible. The experiment lasted about 90 min, including initial briefing, debriefing, calibration, potential recalibrations, and short self-paced breaks between the blocks.

### Data analysis

Recorded gaze data were processed using the SR Research EyeLink (SR Research Ltd., Kanata, Ontario, Canada) blink and saccade detection algorithm. Saccades were detected when the change in recorded gaze location exceeded  $0.1^\circ$ , velocity exceeded  $30^\circ/\text{s}$ , and acceleration exceeded  $8,000^\circ/\text{s}^2$ . Data were pre-processed in MATLAB and statistical tests were run in R.

Eye tracking data from 10,944 trials were analyzed for first saccade’s direction (the fraction of trials in which the first saccade after the continuation of the movies was directed to the target movie). For an analysis of saccadic target RTs, which is the time between the onset of the movies and the initiation of the first saccade that landed on the target movie, we

excluded saccadic RTs outside  $\pm 2.5$  *SD* around the participant’s condition means.

## Results

### First saccade direction

For each participant, the percentage of trials in which the very first saccade went directly to the target movie was analyzed with regard to the within-participant factors cut type, location, and color rendition (see Figure 7A). First, there was a main effect of cut type,  $F(2, 46) = 39.2$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.062$ . The condition in which the lowest fraction of early saccades landed on the target movie was the BSC ( $M = 53.0\%$ ), while a significantly higher proportion of first saccades to the target movie was observed in WSC ( $M = 56.8\%$ ,  $p = 0.003$ ,  $d = 0.19$ ). In BL trials ( $M = 60.5\%$ ), the proportion was significantly higher than in WSC ( $p = 0.003$ ,  $d = -0.19$ ) and BSC ( $p < 0.001$ ,  $d = -0.39$ ). Second, there also was a main effect of location,  $F(1, 23) = 149.7$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.747$ , with more first saccades directed to the target movie when the location repeated ( $M = 77.2\%$ ) than when it switched ( $M = 36.3\%$ ,  $d = 2.07$ ). Finally, a significant interaction of Cut Type  $\times$  Location,  $F(2, 46) = 8.6$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.019$ , and follow-up *t* tests showed that after location repetitions, the fraction of first saccades toward the target movie was significantly lower in BSC ( $M = 71.4\%$ ) than in WSC ( $M = 79.3\%$ ,  $p < 0.001$ ,  $d = -0.98$ ) or in BL trials ( $M = 80.8\%$ ,  $p < 0.001$ ,  $d = -1.02$ ). However, there was

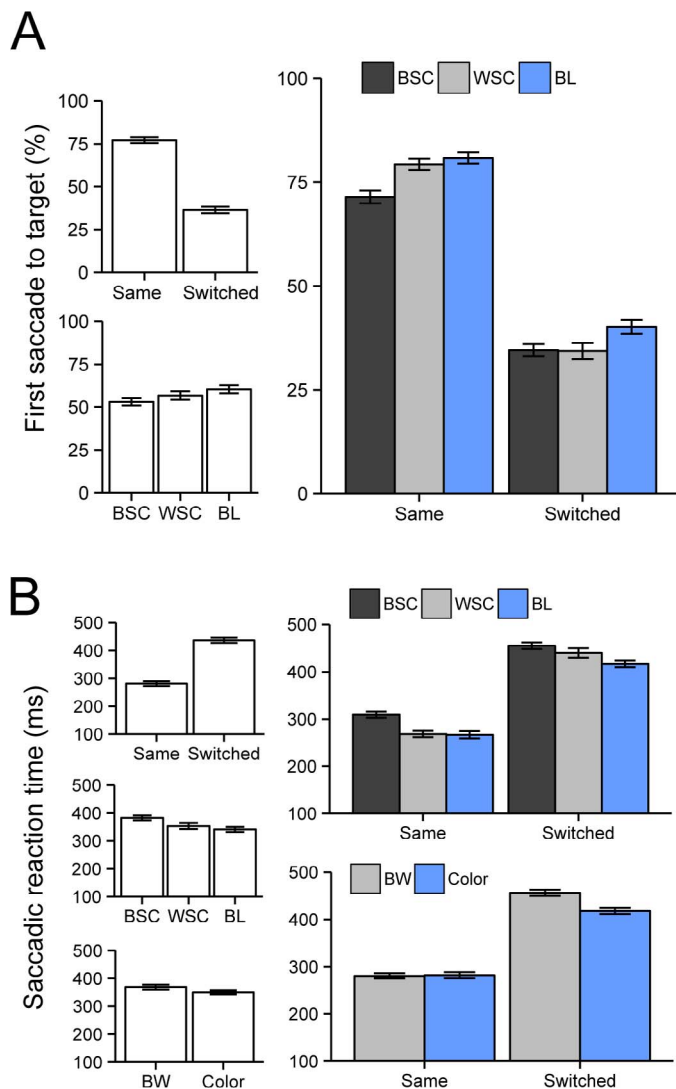


Figure 7. Significant results of the ANOVAs of Experiment 2. (A) Depicted are the proportions of first saccades landing on the target (on the ordinate) as a function of location (top left in A: same; switched), cut type (bottom left in A: BSC, WSC, or BL), and of location and cut type (right in A). (B) Mean saccadic RTs (on the ordinate) as a function of location (top left in B: same; switched), of cut type (center left in B: BSC, WSC, and BL), color (bottom left in B: BW; color), and of location and color (bottom right in B). Error bars represent  $\pm 1$  SEM.

no significant difference in this measure between WSC and BL trials ( $p = 0.580$ ). Conversely, following location switches, the fraction of first saccades going directly to the target was significantly higher in BL trials ( $M = 40.2\%$ ) than in WSC ( $M = 34.3\%$ ,  $p = 0.005$ ,  $d = 0.38$ ) or in BSC ( $M = 34.5\%$ ,  $p < 0.001$ ,  $d = 0.55$ ), while this value did not differ significantly between WSC and BSC. The main effect of color on first saccade direction was not significant,  $F(1, 23) = 2.75$ ,  $p = 0.110$ , and there were also no significant interactions between color and any of the other independent variables. We also repeated the analysis with arcsine-transformed propor-

tions and obtained a qualitatively identical pattern of significant results.

### Saccadic RT

In a second step, we analyzed saccadic target RTs. For this analysis, we excluded 379 trials without saccade toward the target movie or with saccadic target RT outside  $\pm 2.5$  SD around the participant's condition mean (with 96.5% of the dataset remaining). Significant effects are depicted in Figure 7B. First, a significant main effect of cut type,  $F(2, 46) = 29.1$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.044$ , and post hoc tests showed that saccadic RT in BSC ( $M = 382$  ms) was significantly longer than in WSC ( $M = 354$  ms,  $p < 0.001$ ,  $d = -0.30$ ) or in BL trials ( $M = 342$  ms,  $p < 0.001$ ,  $d = -0.48$ ), whereas saccadic RT was not significantly different in WSC and BL trials ( $p = 0.110$ ). Second, there was a main effect of location,  $F(1, 23) = 108.0$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.492$ , with shorter saccadic RTs when location was repeated ( $M = 281$  ms) than when it switched ( $M = 437$  ms,  $d = 1.70$ ). Third, there was a main effect of color rendition,  $F(1, 23) = 13.9$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.013$ , with shorter saccadic RTs in color ( $M = 350$  ms) than in BW ( $M = 369$  ms,  $d = 0.20$ ) movies. Fourth, an interaction of Color  $\times$  Location,  $F(1, 23) = 9.7$ ,  $p = 0.005$ ,  $\eta_G^2 = 0.015$ , and follow-up  $t$  tests showed that the beneficial effect of color on saccadic RTs was present following location switches ( $\Delta RT = 38$  ms,  $p < 0.001$ ,  $d = 0.56$ ), but not when the location repeated ( $\Delta RT = 1$  ms,  $p = 0.820$ ,  $d = 0.03$ ). Fifth, a significant interaction of Cut Type  $\times$  Location,  $F(2, 46) = 4.3$ ,  $p = 0.020$ ,  $\eta_G^2 = 0.005$ , and follow-up pairwise comparisons for cut type on each level of the location variable showed that when locations repeated, saccadic RT was significantly longer for BSC than for WSC ( $\Delta RT = 41$  ms,  $p < 0.001$ ,  $d = 0.88$ ) or BL trials ( $\Delta RT = 43$  ms,  $p < 0.001$ ,  $d = 0.77$ ), while there was no significant difference between WSC and BL trials ( $\Delta RT = 2$  ms). Conversely, when locations switched, saccadic RTs in BSC were significantly longer than in BL trials ( $\Delta RT = 38$  ms,  $p < 0.001$ ,  $d = 0.76$ ), but the difference between BSC and WSC did not reach significance ( $\Delta RT = 16$  ms,  $p = 0.256$ ,  $d = 0.20$ ). Yet, saccadic RTs in WSC were still significantly longer than in BL trials ( $\Delta RT = 22.8$  ms,  $p = 0.033$ ,  $d = 0.32$ ).

### Time course of overt attention

In an additional exploratory analysis, we looked at how the participants' attention unfolded over the course of the first second following a cut. Based on the recorded horizontal gaze coordinates (cleaned from blinks and off-screen fixations) we computed the average horizontal gaze coordinates in time bins of 100 ms (see Figures 8 and 9). Within each time bin, we

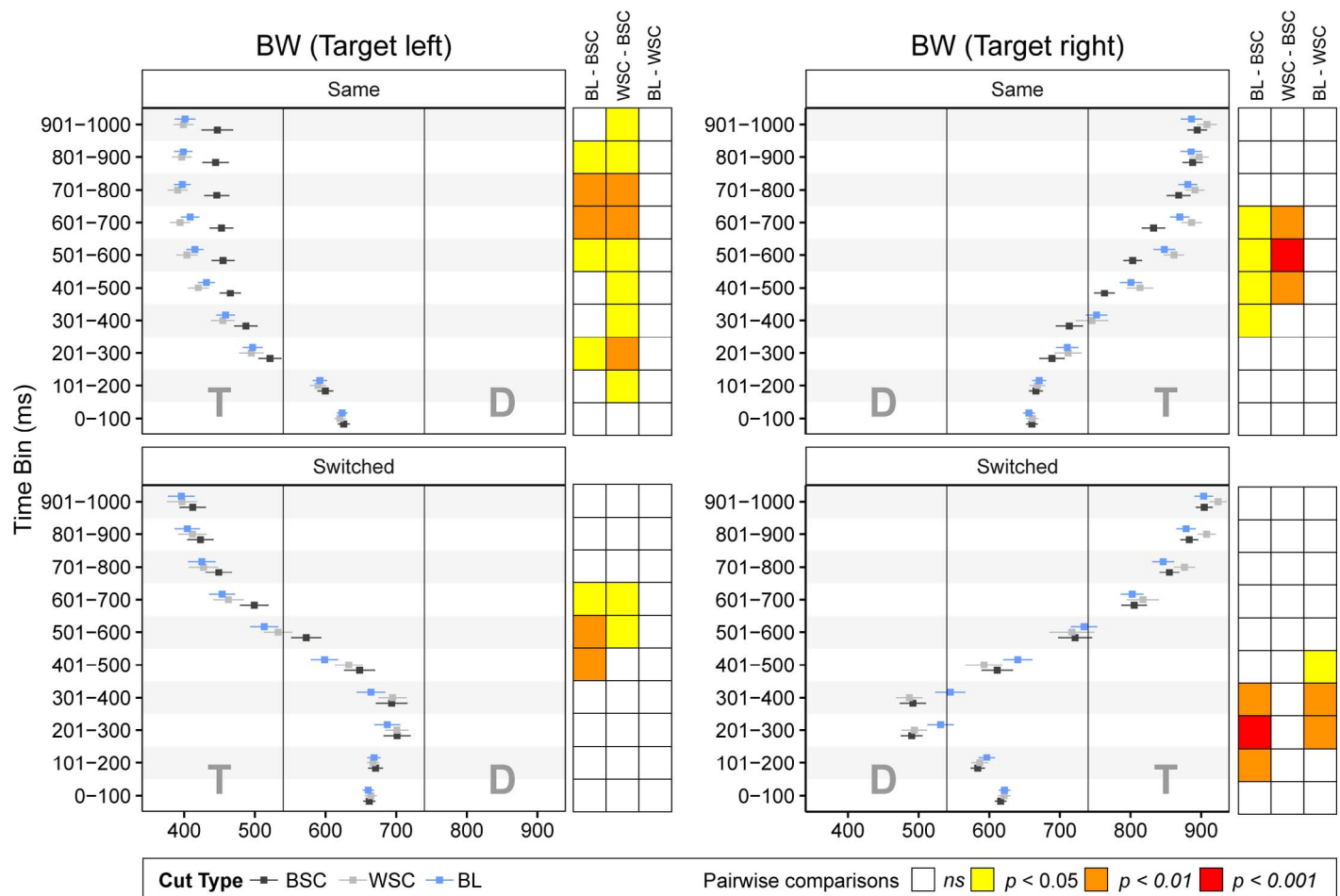


Figure 8. Time course of overt attention in Experiment 2 in trials where the movies were shown in BW. Gaze coordinates are averaged for time bins of 100 ms since trial onset. Left panels show trials in which the target movie (T) was on the left location and the distractor (D) on the right; right panels show the opposite cases. Upper panels show trials in which the location of the target was repeated, lower panels where it switched. Within each time bin, two-sided  $t$  tests (Bonferroni-corrected) were used to mark significant differences between BL, WSC, and BSC. Error bars represent  $\pm 1$  SEM (between participant variance).

checked for significant differences between cut type conditions.

Figure 8 depicts the time course of gaze directions for trials in which the movies were presented in BW. The data shows that the overall higher similarity of BL and WSC resulted in stronger gaze preferences for the target movie relative to the distractor movie, and this tendency was present already early in the trial in some of the display configurations. Relative to BSC, WSC generated similarly pronounced gaze preferences for the target movie as BL cuts (with the exception of trials in which the location of the target switched from left to right).

Figure 9 depicts the complementary data for trials in which the movies were presented in color. Again, relative to BSC, stronger gaze preferences for the target movies were found in the visually more similar BL cuts in all display configurations. WSC led to pronounced gaze preferences for the target movie in displays where the target was presented right, though these differences

were not significant in displays where the target was presented left. As in the BW data, BL and WSC were mostly not different from each other.

### Relationship between visual similarity and saccadic RTs

As in Experiment 1, to check if differences in low-level similarity might have contributed to the observed data we fitted a linear model with the PCA-based similarity index and the location factor as fixed effects, separately to the median saccadic RT data from color and BW conditions. The PCA-based similarity index was derived from a range of typical low-level feature similarity dimensions, computed for the last pre-cut and the first post-cut frame for each cut (see Figure 3, and PCA-based similarity index in the Methods of Experiment 1, for details). The model for color movies yielded a significant main effect of location,  $F(1, 452) = 162.65$ ,  $p < 0.001$ . In addition, the data also yielded a

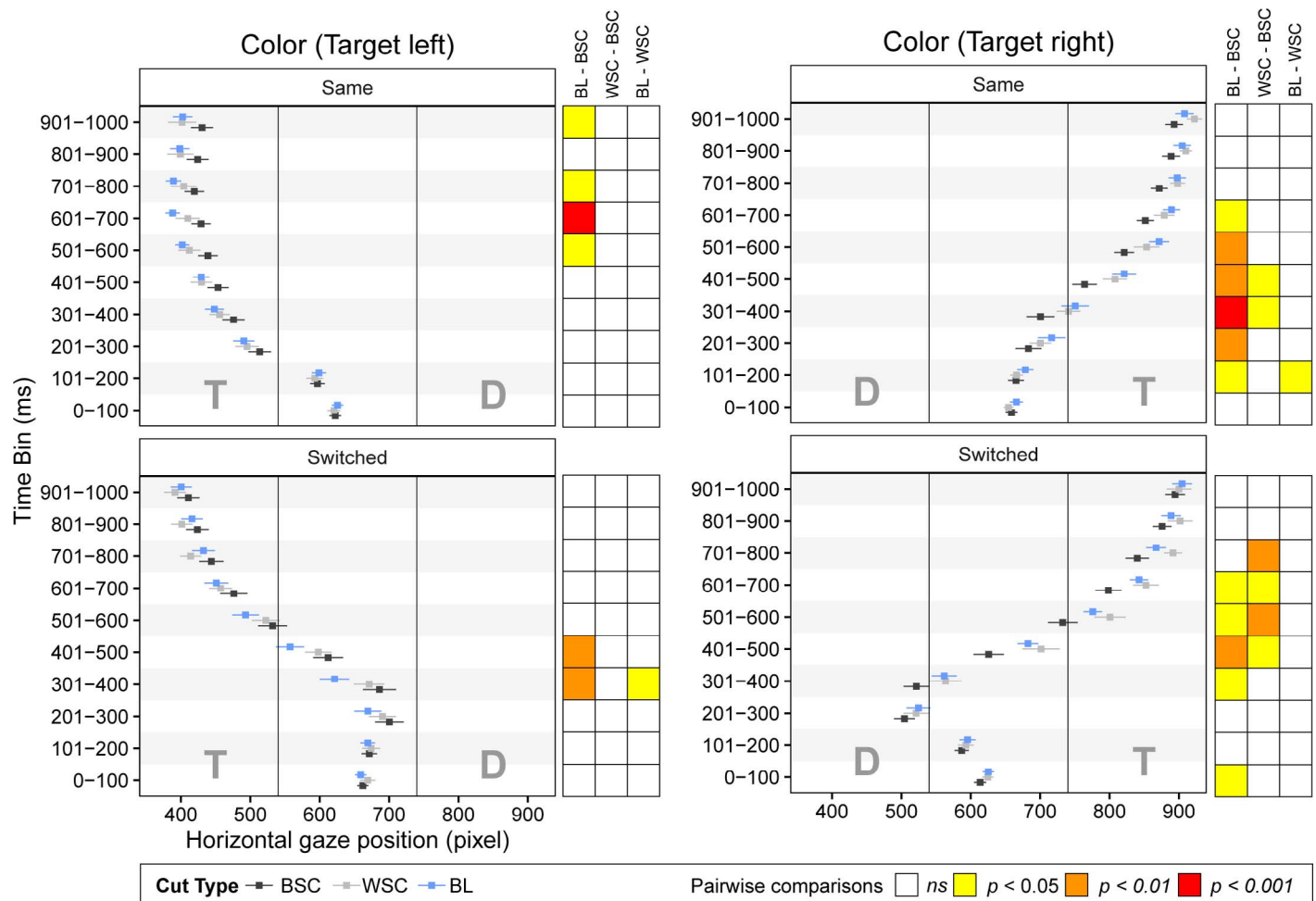


Figure 9. Time course of overt attention in Experiment 2 in trials where the movies were shown in color. Gaze coordinates are averaged for time bins of 100 ms since trial onset. Left panels show trials in which the target movie (T) was on the left and the distractor (D) on the right; right panels show the opposite cases. Upper panels show trials in which the location of the target was repeated, lower panels show trials where its location switched. Within each time bin, two-sided  $t$  tests (Bonferroni-corrected) were used to mark significant differences between BL, WSC, and BSC. Error bars represent  $\pm 1$  SEM (between participant variance).

smaller but significant main effect of similarity,  $F(1, 452) = 4.51$ ,  $p = 0.034$ . The interaction Location  $\times$  Similarity was not significant,  $F(1, 452) = 0.01$ ,  $p = 0.919$ . The model coefficients reflected increased saccadic RTs after location switches,  $b = 134.34$ ,  $SE = 10.53$ ,  $p < 0.001$ , and decreasing saccadic RTs with higher degrees of similarity,  $b = -16.51$ ,  $SE = 7.77$ ,  $p < 0.001$ .

For the BW data, the model again yielded a significant main effect of Location,  $F(1, 452) = 254.31$ ,  $p < 0.001$ . However, as in Experiment 1, the similarity predictor did not show a statistically significant effect,  $F(1, 452) = 3.71$ ,  $p = 0.055$ , and neither did the interaction of Location  $\times$  Similarity,  $F(1, 452) = 0.02$ ,  $p = 0.902$ . The model coefficients again reflected increased RTs after location switches,  $b = 171.59$ ,  $SE = 10.76$ ,  $p < 0.001$ , and, although not statistically

significant, decreasing RTs with higher degrees of similarity,  $b = -15.29$ ,  $SE = 7.94$ ,  $p = 0.055$ .

Figure 10 depicts the median correct saccadic RTs per cut as a function of the degree of visual similarity, separately for the location and the color manipulations, with a linear fit on the subsets of WSC, BSC, and BL conditions. Table 4 depicts the correlations between the median saccadic RTs and visual similarity for each of the different conditions. These results show that significant negative correlations between similarity and saccadic RTs were observed for WSC, after location switches. In the other conditions of BSC and WSC, we found only nonsignificant negative relationships. In the BL condition, which already contained the highest similarity scores, the remaining variance in similarity scores did not systematically correlate with median saccadic RTs.

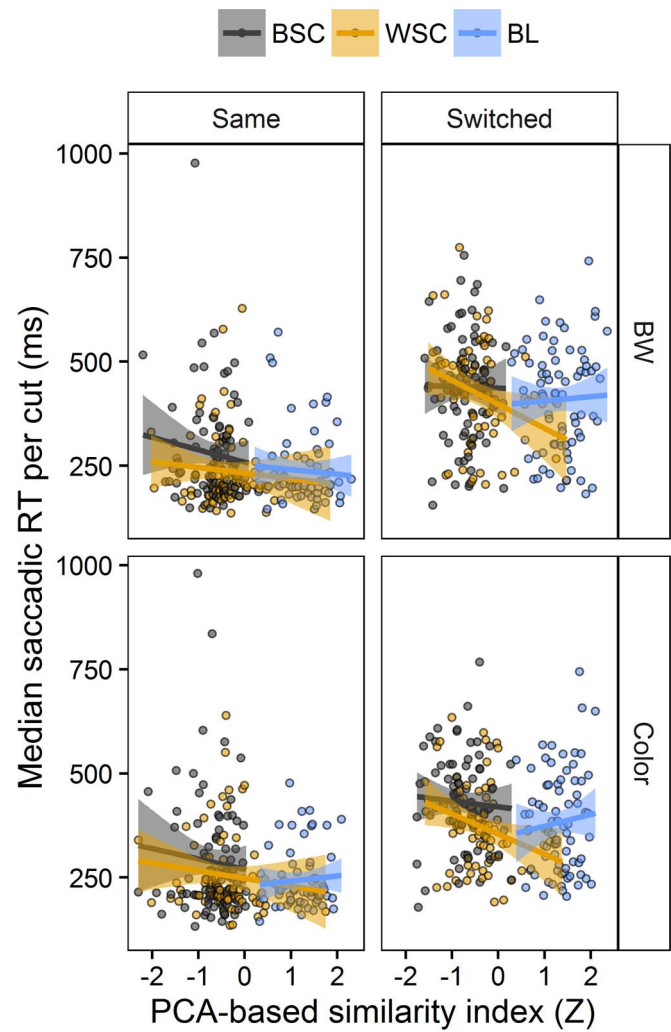


Figure 10. Median saccadic RT per cut (from 24 subjects) in Experiment 2 as a function a PCA-based similarity index that represents how visually similar the pre- and the postcut image are. Data are plotted in separate panels for the location and the color factors. Within each panel, colored lines and shaded surfaces ( $\pm 1$  SE) depict a linear fit for subsets of the data according to cut type (BSC, WSC, or BL).

Discussion

Experiment 2 supported the conclusions of Experiment 1 regarding a facilitative effect of visual feature similarity for the deployment of attention after cuts. Again, we instructed participants to actively follow an edited movie on a particular topic, and we evaluated how quickly they would be able to shift their gaze back to the target movie continuations following cuts depending on the degree of visual similarity. As predicted, gaze shifts to the target movie were initiated faster and a larger proportion of the first saccades landed directly on the target movie when the visual similarity across the cut was high (as in BL or

	BSC	WSC	BL
BW			
Same	−0.11	−0.08	−0.07
Switched	−0.01	−0.26*	0.04
Color			
Same	−0.08	−0.11	0.07
Switched	−0.06	−0.26*	0.09

Table 4. Pearson correlation coefficients and  $p$  values for the relationship of similarity and reaction time, as a function of cut type (BSC/WSC/BL), color versus black and white, and location (same/switched). *Note:* BSC = between-scene cuts; WSC = within-scene cuts; BL = baseline; BW = black and white. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ .

WSC) than when it was low (as in BSC), although these benefits differed slightly between the measures of first saccades and saccadic RTs, depending on whether or not the location of the movies stayed the same.

As in Experiment 1, we also obtained a strong spatial continuity effect: Viewers showed a strong bias of directing their gaze to the location that was previously occupied by the target movie, and they avoided the location that was previously occupied by the distractor movie. One possible explanation for this relatively strong bias toward the previous location could be that although participants were informed that movies will frequently switch locations after cuts, they might not have assumed a completely equal probability for location switches on each trial. To note, returning the gaze to the recently fixated location—even when it was occupied by the distractor movie—and then switching to the alternative location of the target movie was not a very costly strategy as erroneous saccades to the distractor did not lead to any negative feedback. However, the bias for the last fixated location could also have reflected spatial priming of attention, which is a common finding in visual search experiments with simple static object arrays (Kristjánsson & Campana, 2010; Maljkovic & Nakayama, 1996). Similarly, experiments on gaze behavior in static scenes revealed that humans frequently make saccades toward recently fixated scene locations, a phenomenon called “facilitation of return” (Smith & Henderson, 2009). This effect can be particularly strong for saccades toward sudden onsets at recently fixated locations, possibly reflecting an oculomotor mechanism that supports processing of visually complex scenes, as any visual change at a recently fixated location implies that information encoded from that location is no longer valid and needs to be updated (Smith & Henderson, 2009). In our experiment, the onset of the movies after each cut could have triggered return saccades to the previously fixated target locations, which would argue against an explicit attentional strategy of the participants to attend to the

last location, and could explain why the benefit for repeated locations was even more pronounced in Experiment 2 than in Experiment 1.

Within saccadic RTs, we also replicated the color benefit and its interaction with location. Color again benefited attention shifts to the target movies independently of cut type, and again this effect was stronger after location switches than repetitions. One possible origin of this interaction is that there might have simply been more space for the color effect in the location-switch conditions because the tendency to attend to the previous target location could have reflected a bias operating independently of and before taking the perceptual evidence into account. For the same reason location repetitions might have undermined effects of a cut's visual similarity.

Taken together, Experiments 1 and 2 clearly demonstrated that when viewers actively follow a particular edited movie, they are able to deploy their visual attention more rapidly and efficiently to postcut scenes when these match the memory of the preceding scene. Yet what remains unclear after both experiments is whether the observed effects of visual similarity reflect an automatic operation of attention during edited movie viewing, in the sense that attention is primed in a stimulus-driven way via the repetition of visual features across the cut (e.g., Kristjánsson & Campana, 2010; Theeuwes, 2013) or whether these effects depend on the fact that participants were actively following the target movie, as this was their actual task. If the effects were due to purely stimulus-driven processes, we should obtain similar effects of WSC on attention if participants would only passively view a single precut scene, followed by two alternative postcut scenes, one of which was a continuation of the precut scene. Experiment 3 tested this possibility.

### Experiment 3: Does the similarity effect persist under passive viewing conditions?

Here, we studied if the observed effects of visual similarity were stimulus-driven (possibly involuntary) or depended on the viewer's task to attentively follow the target movie. To that end, we left away any additional task and created a situation that did not require participants to attentively follow a particular movie. Following each cut, participants were explicitly allowed to look at whichever of the two postcut scenes they found more interesting, and thus they had freedom of choice for viewing either the target movie (here: the continued movie because there was no task

that designated a more relevant target movie) or the distractor movie (here: the postcut clip taken from a different movie than the precut clip). To allow an experimental measurement of the effects of cut type and color on overt attention following cuts, it was necessary to constrain the movie scene participants viewed before each cut. Instead of always showing two movies side by side, we started each trial with only one, centrally presented movie clip. This single movie was followed by its own continuation (in the form of a BSC, WSC, or BL cut) at one of the peripheral locations, and the continuation of another, unrelated movie, at the alternative location. To further minimize the likelihood that participants would build up some kind of intrinsic interest in a particular movie sequence, we presented trials (each consisting of a single precut clip and two alternative postcut clips) in an order that was randomized for each participant across the complete stimulus set. For example, a trial with a BMX racing target clip and a skateboarding distractor clip could be directly superseded by a trial showing a skiing target clip and a surfing distractor clip.

Only if the tendency to deploy attention toward visual similarity was a stimulus-driven effect, we expected to see the same preference as in Experiments 1 and 2 for (a) continued (target) movies over alternative (distractor) movies and (b) among the continued movies a preference for BL and WSC over BSC. In contrast, if the tendency to attend to visual similarities following a cut is under top-down control and primarily facilitates attention shifts across cuts when viewers are actively engaged in a particular movie (which was before ensured by the viewing task in Experiments 1 and 2), we expected not to see much evidence for this attentional pattern in the present experiment.

## Method

### Participants

Twenty-four (13 female) new participants with a mean age of 24 years (range 19–38 years) were recruited from the same undergraduate psychology student population in exchange for partial course credit.

### Movie stimuli

The same movies and cuts as in Experiments 1 and 2 served as stimuli. Also, the apparent size of each movie was the same as in Experiment 2. Yet, to minimize the potential that our participants created an interest in the continuation of a more extended movie, instead of presenting blocks of continuous movies, we presented the cuts in randomized order. This was done because

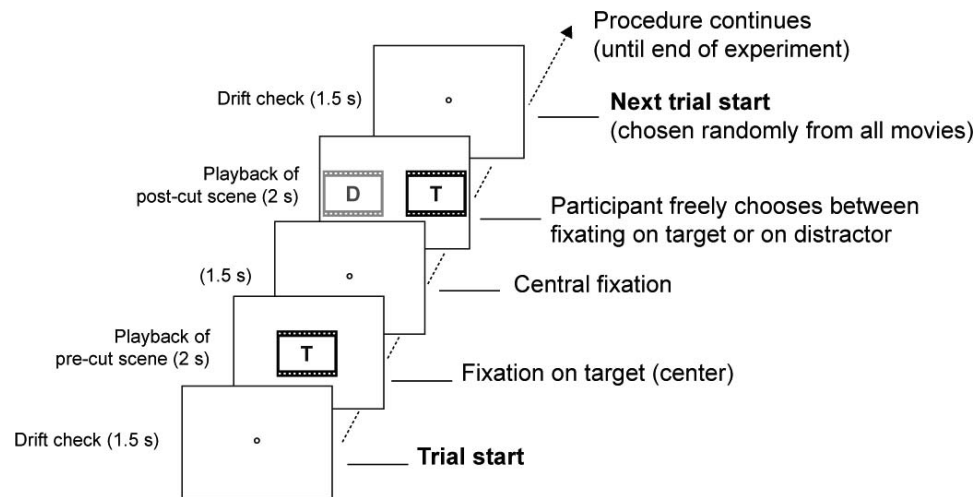


Figure 11. Schematic procedure in Experiment 3. Participants could freely choose if they looked at the continuation (T) of a first centrally presented single shot or whether they looked at an unrelated movie (D). In Experiment 3, movies did not show the same type of sports throughout a block. Instead, the sequences of pre- to postcut shots were picked randomly from the whole set of movies and changed from trial to trial.

with too much interest in the movies, our participants could possibly again develop a top-down interest in a movie and actively search for repetitions in the postcut clips. To present the cuts in randomized order, we sliced each of the movies into a number of short clips corresponding to the 2 s prior to a WSC, BSC, or BL cut, and the two seconds following a WSC, BSC, or BL cut. For the sake of consistency, each short clip was paired with the short clip from the complementary movie of the original set. Hence, the same numbers, types, and pairs of cuts were used as in the previous experiments.

A trial is depicted in Figure 11. Compared to Experiments 1 and 2, the procedure was modified for an assessment of the influence of viewing a particular movie image on attentional orienting after WSC, BSC, or BL cuts under complete free-viewing conditions: First, only the pre-cut shot was presented as a single movie at screen center. After the interruption, this movie clip continued either at the left or at the right location (at an eccentricity of  $7.2^\circ$ ), and simultaneously the continuation of an unrelated clip (taken from the distractor movie in the original set) appeared at the alternative location.

### Apparatus

Eye movements were recorded using the same setup and settings as in Experiment 2.

### Procedure and design

Again, the experiment consisted of 456 trials, which resulted from crossing the variables (a) cut type (BSC vs. WSC vs. BL), and (b) color rendition (color vs.

BW). Location (same vs. switched) was not included as an experimental variable because each trial started with a single movie at screen center and continued with one left and one right postcut clip, so that locations changed in all experimental conditions. Color rendition was counterbalanced across participants. At the start of the experiment, participants were informed that they had to simply attentively look at the movie clips. They were told that one clip at the center was followed by two movie clips side by side. They were explicitly informed that whenever there were two movies on the screen, they could freely choose which one they wanted to view and that they could also switch between the movies, if they wanted. Participants were also asked to fixate the small circle at screen center whenever it was present. Trials were presented in random order, and in four blocks. Between blocks, participants were shown a pause screen and could rest briefly before starting the next block by pressing the enter key on the PC keyboard. Due to the modified procedure, and the aim to measure attention under passive viewing conditions, we omitted any additional task such as digit reporting (Experiment 1) or reporting of target movie location (Experiment 2).

Each trial started with a drift check or a central fixation, followed by a single movie at screen center for 2 s. Next, the central fixation circle reappeared for 1 s, followed by the continuation of the first shot at one of the two peripheral locations and the second half of an unrelated short clip at the alternative location. As soon as the playback of the two movies started, participants freely chose to which movie they directed their eyes. The participants' viewing behavior was evaluated with respect to the likelihood of looking at the continuation versus the novel movie. A complete experimental run

took about 60 min, including participant briefing, debriefing, calibration, potential recalibrations during the experiment, and short self-paced breaks in between the blocks.

### Data analyses

Data were preprocessed in MATLAB and statistical tests were run in R. Saccades were identified using the same settings as in Experiment 2. Eye tracking data were collected from a total of 10,944 trials. We analyzed the participants' viewing behavior after the onset of the two alternative movies with respect to two different dependent variables: (a) first saccade direction, which was the percentage of trials in which the first saccade to either of the movies was directed to the continued (target) movie instead of to the novel (distractor) movie, and (b) saccadic RTs for the first saccade toward the target movie and toward the distractor movie, respectively. Analyses of first saccade direction was based on data from all trials (with the exception of trials in which neither the target nor the distractor movie were fixated). For analyses of saccadic RTs toward the target and the distractor movie, we excluded trials in which the respective movie was not fixated.

## Results

### First saccade direction

Across all conditions, the participants' likelihood for making the first eye movement to the target instead of the distractor movie was not different from chance ( $M = 50.2\%$ ,  $SD = 6.13\%$ ). There were also no significant effects of cut type,  $F(2, 46) = 0.05$ ,  $p = 0.950$ , or color,  $F(1, 23) = 0.41$ ,  $p = 0.520$ , and there was also no interaction of Cut type  $\times$  Color,  $F(2, 46) = 0.25$ ,  $p = 0.777$ .

### Saccadic RT for target and distractor movie

We excluded 1,708 trials (15.6%) in which participants did not make any saccades to the target movie. Further 283 trials (2.6%) were excluded because saccadic RT was outside a range of  $\pm 2.5 SD$  around the participant's individual condition mean. On average, it took participants 447 ms ( $SD = 83.6$  ms) to initiate the first saccade toward the target movie. Again, this value did not differ significantly depending on cut type,  $F(2, 46) = 0.22$ ,  $p = 0.81$ , or color,  $F(1, 23) = 1.99$ ,  $p = 0.17$ .

We ran the same type of analysis for saccadic RTs toward the distractor movie. We excluded 1,731 trials (15.8%) in which participants did not make any saccade to the distractor movie. Another 303 trials (2.7%) were

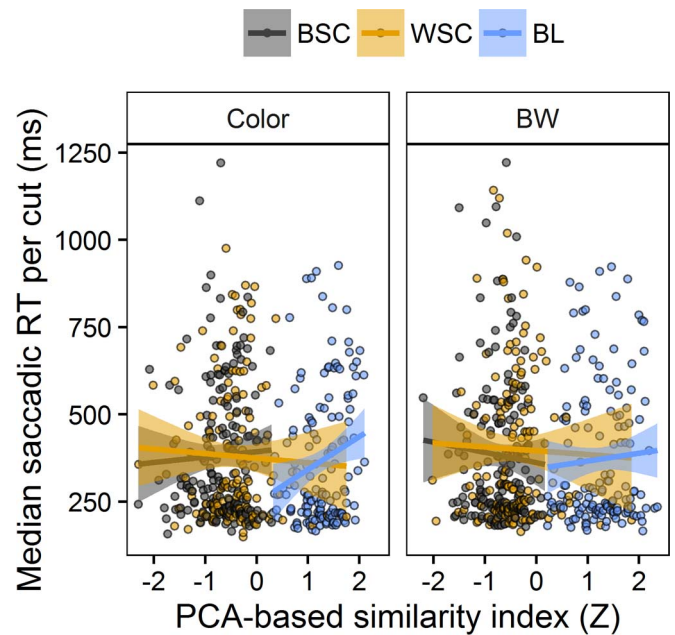


Figure 12. Median saccadic RT per cut (from 24 subjects) in Experiment 3 as a function of a PCA-based similarity index that represents how visually similar the pre- and the postcut image were to one another. Data are plotted in separate panels for the two conditions of the color manipulation. Within each panel, colored lines and shaded surfaces ( $\pm 1 SE$ ) depict a linear fit for subsets of the data according to cut type (BSC, WSC, or BL).

excluded because saccadic RT was outside a range of  $\pm 2.5 SD$  around the participant's individual condition mean. On average, it took participants 442 ms ( $SD = 81.4$  ms) to make the first saccade toward the distractor movie. Here, the analysis yielded a significant main effect of cut type,  $F(2, 46) = 4.79$ ,  $p = 0.013$ ,  $\eta_G^2 = 0.027$ . Saccadic RTs toward the distractor movie were significantly shorter after BL cuts ( $M = 423$  ms) than after BSC ( $M = 453$  ms;  $p = 0.037$ ,  $d = 0.36$  for the difference between the two), whereas the difference between BL and WSC ( $M = 448$  ms) was of a similar size but did not reach significance ( $p = 0.057$ ,  $d = 0.36$ ). As before, there was no significant main effect of color,  $F(1, 23) = 0.14$ ,  $p = 0.71$ .

### Relationship between visual similarity and saccadic RTs

Across all trials, for the color data, (PCA-based) similarity did not have a significant effect,  $b = -0.607$  ( $SE = 9.242$ ),  $p = 0.950$ . Also for the BW data, similarity did not have a significant effect on saccadic RTs,  $b = -7.76$ ,  $SE = 9.88$ ,  $p = 0.44$ . For the sake of consistency, we again plotted and computed median saccadic RTs as a function of the PCA-based similarity index (see Figure 12). Table 5 lists the correlations between visual similarity and saccadic RTs for each of the experimental conditions. Only the BL cuts showed a statistically

	BSC	WSC	BL
BW	−0.06	−0.03	0.06
Color	0.04	−0.04	0.20*

Table 5. Pearson correlation coefficients and *p* values for the relationship of similarity and reaction time, as a function of cut type (BSC/WSC/BL), color versus black and white. *Note:* BSC = between-scene cuts; WSC = within-scene cuts; BL = baseline; BW = black and white. \**p* ≤ 0.05.

significant relationship: A larger degree of visual similarity was positively associated with saccadic RT.

Time course of overt attention

For an overall impression of the participant’s viewing behavior, and to compare it with the results of Experiment 2, we again performed an exploratory

analysis of the average time course of the horizontal gaze position over the course of the first second after the onset of the movies. Horizontal gaze coordinates were averaged in 10 time bins of 100 ms. Within each time bin, we checked for significant differences between the cut type conditions, separately for the BW trials (Figure 13, upper panels) and the color trials (Figure 13, lower panels). Different from Experiment 2, and in line with the main ANOVA of Experiment 3, there were no gaze preferences for the target movie, and there were also no clear differences between the three cut type conditions.

Discussion

After Experiment 3, it is clear that the facilitative effects of visual similarity on the deployment of

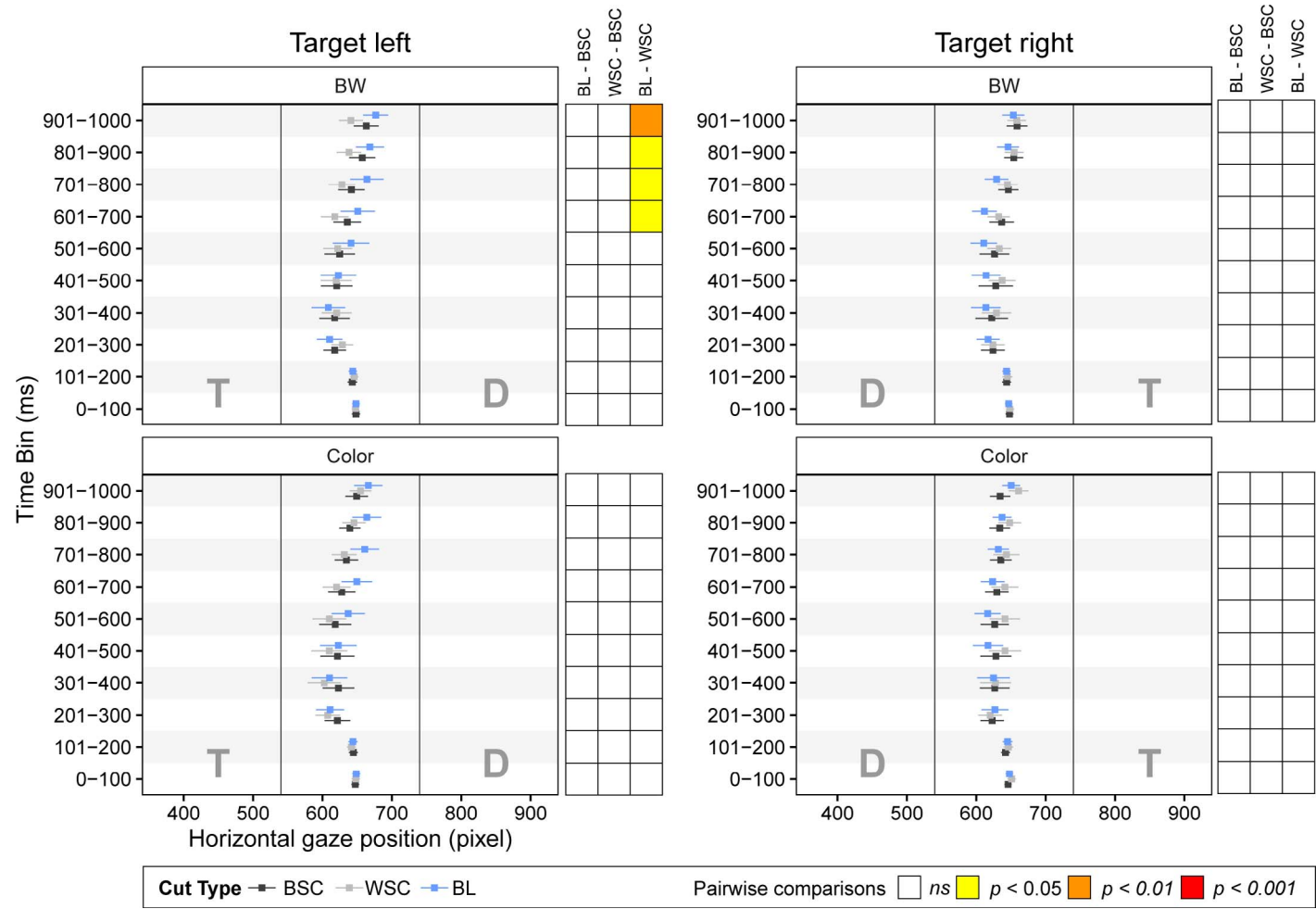


Figure 13. Time course of overt attention in Experiment 3. Gaze coordinates are averaged for time bins of 100 ms since trial onset. Left panels show trials in which the target movie (T) was presented on the left and the distractor (D) on the right; right panels show the opposite cases. Upper panels show trials in which the location of the target was repeated; lower panels show trials where it switched. Within each time bin, two-sided *t* tests (Bonferroni-corrected) were used to mark significant differences between BL, WSC, and BSC. Error bars represent ±1 SEM (between participant variance).

attention to a postcut scene (Experiments 1 and 2) do not reflect purely stimulus-driven processes. It thus appears as if visual similarity helps viewers keep their attention focused on an edited movie when they are actively following it (as we made sure in Experiments 1 and 2 by means of the viewing task)—but that visual similarity across two subsequent clips does not automatically capture viewers' attention when they are only passively viewing a rather random clip sequence (Experiment 3).

Under the passive viewing conditions of Experiment 3, participants did not show a uniform preference for or against continued (target) over novel (distractor) postcut clips. Related, among the postcut target clips, not even following cuts with the highest possible visual similarity (the BL condition), viewers deployed their attention to the target continuation significantly more often than to the distractor. In fact, the only evidence for any preference for one of the two alternative clips was found in distractor-directed saccadic RTs within BL trials: In these trials with a complete continuation of the precut target clip, saccades to the distractor continuations were initiated faster than in trials in which the relation of the pre- to postcut target clip was a BSC. This could suggest that when viewers are confronted with little novel information in a postcut target continuation, they might rather opt for the alternative which is, under these conditions, more interesting because it presents novel information. Such an interpretation would be well in line with research that looked into the capture of attention by surprising or novel visual information (Horstmann, 2002, 2005; Itti & Baldi, 2009). For example, Horstmann (2005) used visual search for targets and showed that his participants swiftly attended to those targets that were most discrepant from the expected visual information. Along similar lines, Itti and Baldi (2009) demonstrated that during movie viewing participants preferentially looked at such locations that contained visual information that deviated most strongly from what was presented until this point in time. However, in the case of dynamic scenes, this novelty- or surprise-based effect has been primarily demonstrated as a principle that is important for attentional selection during an ongoing movie take, but the authors did not investigate the validity of this principle after cuts of varying degrees of visual relatedness while participants actively view thematically related sequences of shots.

In conclusion, a crucial variable in determining whether visual similarity affects how viewers direct (or are able to direct) their attention to a postcut scene seems to be the degree to which they are actively following a particular movie. While our viewing task in Experiments 1 and 2 might seem unnatural at first glance, its main advantage is to ensure that all viewers were interested in the target movie. During everyday

movie viewing, of course the task of having to match thematically related scene content across cuts would not be explicit, as there usually is no distractor movie that needs to be discriminated from the movie that is currently being followed. Yet also during everyday movie viewing, participants have to master the implicit same task of relating successive movie images to one another. But what is also clear is that for the viewers to pick up on this task, they need to be interested in what they are watching.

## General discussion

The main goal of the current study was to test if visual feature similarities facilitate attention shifts across cuts in edited dynamic scenes. Before we enter into a more detailed discussion, we briefly summarize the most important novel findings of the present study. Here, we reported at least the following three new major findings. First of all, in Experiment 1, we showed that visual feature similarity across cuts of edited dynamic scenes can facilitate a viewers' covert shifting of attention to the most similar postcut content during movie viewing. This is an important finding as hitherto only influences of cut similarity on overtly conducted eye movements have been demonstrated and because eye movements and covert shifts of attention can sometimes be dissociated (Posner, 1980). According to our Experiment 1, visual similarity across a cut could indeed benefit film comprehension across cuts by allowing a shift of attention to the most similar postcut content and benefiting the selection and recognition of novel information from these attentionally selected positions. Second, we were also able to demonstrate that the facilitative influence of visual similarity on attention could indeed be responsible for some of the facilitated comprehension during film viewing as we also confirmed that visual similarity is indeed higher for one type of continuous cuts, WSC, as compared to a type of discontinuous cuts, BSC. Finally, the current study is the first one that analyzed how the attentional effect of visual cut similarity unfolds across time in the postcut images. We were able to show that the beneficial effects of visual attention were present from as early as 101–200 ms into the postcut frames. This finding is also important in that it shows how quickly the influence of visual similarity on attention in the postcut images can take effect, further supporting that similarity-driven attention could indeed be an important factor in across-cut film comprehension. Of course, all these findings require some qualifications and in-depth discussion, which we do next.

In Experiments 1 and 2, we compared attention shifts to a target movie between visually more similar

WSC (in which the major agent or object did not change across the cut) and less similar BSC (in which the major agent or object changed across the cut). Compared to BSC, WSC (as well as BL trials, in which the postcut image was a direct continuation of the precut image) facilitated both covert and overt attention shifts to the postcut clips when participants were actively following the target movie. Even within these cut type conditions, visual similarity differences were correlated with the participants' (saccadic) RTs. This was true for both BSC and WSC in Experiment 1 and then replicated for WSC in Experiment 2. With the final Experiment 3, we determined if the attentional benefits of feature similarity critically depended on the viewers' attentional engagement with a particular target movie. When the viewers were not actively engaged with a particular target movie, visual similarity did not facilitate attention shifts to a movies' continuation after cuts. Taken together, the results suggest that visual similarities facilitate attention shifts across cuts, but this effect critically depends on the viewer's top-down goal of establishing perceptual relations across subsequent shots.

Besides these effects of major interest, we also found reliable influences of two further variables on attention. First, to determine if the presence of color features was particularly critical for the attentional benefits of visual similarity, we presented half of the movies in color and the other half in BW to our participants. We found that the presence of color features accelerated both manual and saccadic RTs in Experiments 1 and 2, but this effect was independent of the cut type manipulation. While participants were, across all cut type conditions, faster in determining which of the two simultaneously presented videos was the target, the attentional benefits of WSC (and BL cuts) persisted (to roughly the same degree) even when movies were presented in BW. Second, attention shifts benefited from repeating the target movie's location from before the cut. After location switches, the viewers' attention appeared to be biased for the location where the target movie had been displayed in the previous shot. While this location effect was already strong in manual RTs (Experiment 1) it became even more evident in the eye movement data (Experiment 2). Interestingly, deploying attention to the new location after a switch was again facilitated by higher visual similarities (in BL or WSC) on the one hand, as well as by the presence of color features on the other hand.

The results of our study broaden the current understanding of how the human visual system deals with edited material in a wider range of contexts. At the outset, we have mentioned that *continuity editing* techniques (Bordwell & Thompson, 2001) are rather successful in making shot transitions as smooth as possible, sometimes even rendering them subjectively

invisible (Shimamura et al., 2014; Smith & Henderson, 2008; Smith & Martin-Portugues Santacreu, in press). Several authors suggested that processes of attention and working memory are important for perceiving continuity in edited material. However, theoretical conceptions in this area mostly focused on conceptual processes. For example, it was suggested that cinematic continuity depends on the viewers' general expectations about how the movie's narrative will develop (which might in turn influence the viewers' attentional settings; cf. Smith, 2012), as well as relatively coarse working memory representations that rely on the viewers' conceptual interpretation and understanding of the ongoing events (Berliner, & Cohen, 2011; Magliano & Zacks, 2011). We believe that such processes are important for perceiving edited material, especially if it concerns material that contains an extended and more complex narrative structure (which was not the case with our relatively simplistic and silent sports videos). However, our results illustrate that basic perceptual factors, such as the visual similarity between successive shots or the repetition of relevant visual content at a previously attended location have a reliable influence on how quickly the viewers' attention can pick up scene contents that are conceptually related to the preceding shots. As these early attentional selection processes are the necessary precursor for any higher level conceptual interpretation of the scene content, it is important to consider them in order to fully understand how humans perceive edited material.

Our findings can also be related to current theories of the cognitive processes underlying the perception of cinematic continuity. The *attentional theory of cinematic continuity* (Smith, 2012) emphasizes an active role of the individual viewer for perceiving continuity. Our data strongly supports this assumption. If viewers do not have the goal to actively follow a movie, their attention will not be automatically captured by visually or conceptually related content (as it was the case in our Experiment 3). The theory also suggests that the repetition of (screen) location of an attended object is important because it satisfies the viewers' general expectation of spatiotemporal continuity (Smith, 2012, p. 15). In our Experiments 1 and 2, we indeed found that location repetition strongly facilitated attention shifts to the target movie's continuation, which could be interpreted as evidence for this prediction. However, our present findings go beyond this form of simple or content-independent spatial memory effect, and show that the viewers' expectations about the movie's continuation rely (to certain degree) on concrete visual memory of the perceived content. Such memory representations could be instantiated as templates to guide the viewers' attention after cuts. This would allow viewers to immediately refocus their attention on those objects that repeat across cuts, without the need

of a full perceptual analysis of the novel view on the scene, in line with Smith's (2012) conception of *a priori continuity*. Conversely, if the visual relations across shots are not immediately available, viewers would need to engage in a more reflective and interpretative process to understand how the subsequent shots relate to each other, and establish continuity *a posteriori*. This might have been the case in our BSC, where the recognition of a novel scene as belonging to the same category of sports as the previous scene might have taken more time than the direct mapping of repeated visual content across the cut.

Related to this, our present findings regarding visual similarity are also relevant for further research on *event segmentation theory* (Magliano & Zacks, 2011; Zacks & Swallow, 2007; Zacks et al., 2010). According to this theory, viewers establish and maintain a simplified mental model of the currently unfolding situation in working memory, which is considered key to understanding the overall narrative. In WSC, viewers would only need to update an established model of the situation after each cut, as long as they immediately recognize subsequent shots as belonging to the same perceptual event. However, if the features unexpectedly change (as in the case of BSC), viewers would need to set up a new mental model of the changed situation, which is in turn perceived as the onset of new event (for a review, see e.g., Magliano & Zacks, 2011). Based on our findings, we suggest that visual feature similarity is an important moderator variable in this perceptual parsing process. The facilitative effect of visual similarity on the recognition of movie continuations and the facilitated attention shifts to visually related content that we found in our present study could be characteristic of cuts in which viewers do not need to establish a novel mental model of a scene. Smaller visual feature changes at WSC would allow viewers to maintain their current working memory model of the situation. Also in line with a role of visual features, Magliano and Zacks (2011) studied the consequences of cuts with different degrees of continuity on behavioral event segmentation and brain activity using functional magnetic resonance imaging. They found that only BSC elicited a strong preference to report the onset of a new event. Their study also included cuts in which the central agent (i.e., a kid with a red balloon) repeated across the cut, but the scene background changed. Interestingly, these cuts elicited much fewer reports of event boundaries than BSC that did not preserve the focal object. Moreover, there were minor differences to actual continuity cuts, in which the scene background was preserved along with the central agent. This was true even though the pattern of brain activity was quite different between these conditions. The authors suggested that higher level conceptual processes might be involved in bridging the low-level discontinuities across scene background

changes (Magliano & Zacks, 2011). Based on our present results, we suggest that visual similarity of the dominant foreground scene content contributes to maintaining continuity despite scene background changes. Importantly, visual similarity across cuts does not need to be perfect, as long as the centrally relevant visual features are present in both successive shots. Our WSC, although their visual similarity was not as high as in the BL condition, had similar facilitative effects on attention. Moreover, these benefits reached the BL level in WCSs with particularly high visual similarity scores (see Figures 5 and 10).

## Visual similarity versus semantic continuity

In respect to their major conceptual characteristics, our WCSs were comparable to cuts that could be found in any edited media that follow the continuity editing principles. They established spatiotemporal and action continuity across the cut, which are key ingredients of cinematic continuity (Bordwell & Thompson, 2001; Magliano & Zacks, 2011). However, these properties were inherently related to a stronger continuity of visual features across the cut, and this higher feature similarity distinguished WSC from BSC in our set of stimuli. In contrast, the conceptual (and to some degree even action) relationships were preserved in our WSC and BSC alike (for instance, with a cut from one snowboarder in a red pullover to a different snowboarder in a blue overall). Hence, we argue that the attentional benefits of WSC (including BL cuts) in our data were indeed mediated by the higher visual feature similarity relative to BSC, and not (or at least to a lesser degree) by differences in the conceptual relatedness or the viewers' cognitive understanding of a narrative. It is possible that match-action cuts, which are among the oldest and most powerful continuity editing techniques (Salt, 2009), also capitalize on the visual similarities across cuts. In line with this interpretation, Shimamura et al. (2014) reported data showing that viewers perceive match-action edits as particularly continuous if the depicted onset of an action (in the first shot) and the depicted continuation of the action from a changed perspective (in the second shot) overlap by around 125 ms such that part of the movement from the first shot is repeated in the second shot. The subjective perception of continuity of the temporally overlapping action depictions persisted even when authors inserted a pattern mask between the pre- and the postcut image, suggesting that visual similarities were important for perceptually integrating the different viewpoints of the action. In another study, Smith and Martin-Portugues Santacreu (in press) recently found that following match-action edits, the viewers' overt attention appears to be captured by the motion of the central agent that

repeats across the cut. This might prevent viewers from immediately looking at novel scene content (and instead refocus their attention on the previously attended content) and explain why viewers are often unaware of match-action edits (cf. Smith & Henderson, 2008). In light of our present findings, we would argue that the attentional and perceptual benefits of match-action cuts are promoted by the continuity and similarity of the visual features across the cut.

We want to stress that while our present study highlighted the influence of visual similarities on attention during viewing of edited material, we believe that other more conceptual factors, for instance, the nonvisual memory for subevents within an episode, can also play a decisive role for how viewers drive attention across the cut (Cutting et al., 2012; Magliano & Zacks, 2011; Schwan & Ildirar, 2010; Zacks & Swallow, 2007). In particular, if visual similarities across cuts are low, factors beyond visual features, such as inferred goals (of an actor), could jointly be used to connect visual images. Moreover, although our study is motivated by the aim to understand the factors that determine how humans perceive continuity in edited material, filmmakers have a large repertoire for establishing a subjective sense of continuity despite relatively strong visual discontinuities. Examples are shot-reverse-shot sequences documenting a conversation of two different actors, cross-cutting between different scenes to induce a sense of temporal continuity, or the use of soundtracks to bridge perceptual discontinuities in low-level features (e.g., Arijon, 1976; Bordwell & Thompson, 2001; Smith, 2012). However, our present study, which aimed at isolating visual factors, does not leave much room for semantic effects over and above the influence of visual similarity. It was our intention to use BSC with a smaller visual similarity than WSC and BL cuts. To assess if semantic continuity could facilitate attentional selection over and above the influence of visual similarity, future research could adapt our experimental procedure and measure attentional selection in a set of movies in which the cut conditions are orthogonal with regard to the factors (a) visual similarity and (b) semantic continuity. The visual similarity could be matched between conditions using our quantitative approach. In addition, our experimental procedures might prove useful for testing intermodal effects on attentional selection, such as the role of off-screen sounds for announcing the content of an upcoming shot (c.f., Smith, 2012).

### Bottom-up versus top-down influences on attention in edited dynamic scenes

How do the present findings relate to the existing evidence regarding the factors that determine the

viewers' attention in movies? Much of the prior research has concentrated on assessing the correlations between gaze direction in movies and physical salience, such as contrasts in low-level feature dimensions (e.g., color, orientation, luminance, and motion; e.g., Açıık, Bartel, & König, 2014; Böhme, Dorr, Krause, Martinetz, & Barth, 2006; Carmi & Itti, 2006b; Le Meur, Le Callet, & Barba, 2007; Mital, Smith, Hill, & Henderson, 2011). Other studies have explored potential top-down contributions to attentional and gaze control in movies, for example, by assessing the degree to which gaze direction varies across participants (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Dorr, Vig, & Barth, 2012; Goldstein, Woods, & Peli, 2007; Smith & Mital, 2013) or via interactions of auditory and visual information (e.g., Ross & Kowler, 2013; Vö, Smith, Mital, & Henderson, 2012). Finally, a few studies have investigated how attention and eye movements are influenced by individual perceptual history (Carmi & Itti, 2006a; Dorr et al., 2010; Itti & Baldi, 2009; Wang, Freeman, Merriam, Hasson, & Heeger, 2012).

Together, these previous studies have mainly aimed at understanding the factors that determine gaze behavior *during* a spatiotemporally continuous movie sequence (i.e., a shot). Even though these studies often used edited material, they did not explicitly evaluate attention with respect to the degree of continuity across two consecutive shots. Take the example of Itti and Baldi (2009). In that study, the authors found a preference for novel rather than repeated information. Because these authors computed a Bayesian prior for novelty detection on the basis of a gliding window across all movie images and due to the relative infrequency of cuts, the analysis overrepresented gaze behavior during continuous shots and underrepresented the effect of shot-by-shot relations. Moreover, the particular type of cuts that researchers include in a study could definitely make a difference. Imagine a movie consisting of cuts by which only unrelated material is juxtaposed. In this situation, visual resemblances between pre- and postcut images could generally be low (Carmi & Itti, 2006b), so that the presently found feature similarity influences would never be observed. That might explain why some of the previous studies concluded that bottom-up factors can best explain how viewers allocate their attention after cuts (Carmi & Itti, 2006a; Smith & Mital, 2013). In the absence of the important comparison condition of cuts within scenes, conclusions about the general absence of top-down influences after cuts are premature because the majority of cuts in movies and other edited media occurs between visually and/or conceptually related shots (Smith, 2012).

However, we agree that top-down influences on attention might indeed not always become evident in overt gaze behavior. For example, Loschky, Larson,

Magliano, and Smith (2015) showed that viewers' explicit knowledge and expectations about how the events in a film scene will unfold can sometimes have very little influence on where viewers fixate in a Hollywood film scene. Importantly, professionally produced movies or Hollywood-like material generally minimizes interobserver variability in gaze direction, and often elicits pronounced tendencies to look at the center of the scene (Dorr et al., 2010; Goldstein et al., 2007; Smith & Mital, 2013). Independent of gaze direction, movie viewers might often shift their attention covertly—that is, without moving their eyes (Smith, 2012). Therefore, the spatial distribution of eye fixations in professionally produced content might not always be the sole best indicator of where in the scene viewers actually deploy their attention. Because of these difficulties, we developed our target/distractor paradigm, in which we can measure overt and covert attentional deployment to one of two peripheral spatial locations, irrespective of visual particularities of a certain class of movie stimuli. We believe that this approach can uncover attentional effects that might otherwise go undetected if only fixations within a single full-screen movie are evaluated. Admittedly, this procedure has also its drawbacks. This viewing situation is not very “ecological” in that typically a single movie is viewed at a fixed, known location. Hence, it is possible that aspects of our experimental protocol fostered behavior that would otherwise not be observed. For example, maybe the use of a concomitant distractor movie encouraged a reliance on visual similarities in the participants that would otherwise not be as strong. Yet, data from our lab shows that the preference for scene content that is visually related across cuts also shows up in eye fixations in single full-screen videos, as long as these are visually complex enough to avoid strong center biases. When the view on a particular scene shifts across the cut, the viewers' gaze clusters in image regions that overlap with the pre-cut shot within a few hundred milliseconds after the cut—and this seems to allow viewers to establish the perceptual relations across the cut (Valuch, Seywerth, König, & Ansorge, 2015). A similar tendency to fixate on visual content that repeats across different views on the same scene is also evident from visual memory experiments with static natural scene images (Valuch, Becker, & Ansorge, 2013).

It is important to mention that the attentional effects we observed depended on the viewing task that we gave our participants. In both studies, the task motivated participants to attentively follow the target movie and ignore the distractor movie. Note that the tasks that we employed in Experiments 1 and 2 aimed at creating a viewing situation fairly similar to a more ecological situation where viewers are actively engaged in following a sequence of thematically related movie

shots. Current theories of cinematic continuity generally accept the notion that viewers often need to match the preserved features, or objects, across different views on the same scene (Magliano & Zacks, 2011; Smith, 2012). In other words, to understand an edited movie, viewers need to mentally relate successive shots to one another, and visual recognition is a key process in this viewing situation (e.g., Ansorge et al., 2014; Hochberg & Brooks, 1996). Importantly, prior research shows that visual recognition is supported by the presence of visual similarities between scenes, images, or objects (Brady, Konkle, Alvarez, & Oliva, 2008; Snodgrass & Vanderwart, 1980; Tarr & Bülthoff, 1998). In a standard movie viewing situation, where viewers are engaged with a movie, they actively search for visual information that matches their expectations and helps them comprehend the narrative (Hochberg & Brooks, 1996; Smith, 2012). In contrast, if viewers are not actively engaged in a movie, there would be no benefit in orienting their attention to content that matches their expectations and perceptual hypotheses. This is what we witnessed under the rather passive viewing conditions of Experiment 3: There was no specific task which demanded relating the postcut images to the pre-cut images. Moreover, the movie clips were presented as two-shot sequences only, such that the topics (i.e., the types of sports) and narratives (i.e., the types of actions) that participants saw changed randomly from trial to trial. This means that there was little chance for participants to develop an interest in a sequence of related images belonging to the same topic or narrative. With professionally edited material, it is the job of filmmakers and editors to guide and focus the viewers' attention on the unfolding narrative (Cutting et al., 2010; Hasson, Landesman, Knappmeyer, Valines, Rubin, & Heeger, 2008; Smith, 2012). With our simple stimuli that were largely free from any narrative elements, we decided to make this attentive engagement an explicit task, in order to keep our participants focused on the target movie. However, such a task (whether implicit, as in everyday movie viewing, or explicit, as in our experiments) could be a critical side condition for unveiling any beneficial effects of visual feature similarity on the deployment of attention to the continuation of a movie after a cut.

### Which kind of memory underlies the attentional benefits of visual similarity?

Attentional benefits by repetitions of visual properties are well known from experiments that normally use rather simple and static visual stimuli. Such effects have been discussed under names such as *intertrial priming* of attention (Awh, Belopolsky, & Theeuwes, 2012; Kristjánsson & Campana, 2010; Lamy & Kristjánsson,

2013; Maljkovic & Nakayama, 1994, 1996), *contextual cueing* (Chun & Jiang, 1998; Chun & Nakayama, 2000; Hutchinson & Turk-Browne, 2012), or effects of *visual working memory* on attention (Olivers et al., 2006). To give an example, visual search experiments often consist of several hundred trials in which viewers repeatedly search for and respond to a predefined target object in a simplified visual scene of irrelevant distractors. Performance during such consecutive trials is not independent from one another: In each trial, the viewers' attention is biased toward objects that have the same color, location, or other features in common with the objects that were attended to in a preceding trial (Maljkovic & Nakayama, 1994, 1996). Because viewers are usually not aware of these effects, they have been conceptualized as implicit memory processes (for reviews see Awh et al., 2012; Kristjánsson & Campana, 2010; Lamy & Kristjánsson, 2013).

One important question that arose from research on priming of attention during visual search was to which degree similarity-driven deployment of attention could result from automatic, stimulus-driven (or bottom-up, hence task-independent) priming processes (Theeuwes, 2013), or, alternatively, primarily reflected the viewer's current intentions, in the form of a top-down attentional set defined by the momentary viewing task (Fecteau, 2007; for reviews see Lamy & Kristjánsson, 2013; Theeuwes, 2013). To a certain degree, intertrial priming of attention by visual similarities depends on a viewer's top-down set: As compared to an unprimed stimulus (that changes its color from one trial to the directly following trial), an intertrial primed stimulus attracts more attention if this stimulus is a task-relevant target (Maljkovic & Nakayama, 1994), but a primed stimulus that is a task-irrelevant distractor attracts less attention, even if it is visually salient (Geyer, Müller, & Krummenacher, 2008; Müller, Geyer, Zehetleitner, & Krummenacher, 2009).

In the context of edited dynamic scenes, task relevance could similarly (co-)determine whether attention is quickly directed toward scene content that repeats across cuts—in other words, whether visual feature similarity across cuts influences attention or not. More specifically, memory-guided attention could be most pertinent when the viewer is actively engaged in following a movie over an extended period of time. This is because understanding a sequence of edited images demands quick decisions about whether and how two images that are separated by a cut relate to one another. For that, the viewer needs to (implicitly) recognize the movie image following a cut as showing or not showing the same situation, location, and action as before the cut (Ansorge et al., 2014; Hochberg & Brooks, 1996). However, if a viewer is not actively following a movie, there would be no need to determine how a new shot following a cut relates to the previous

one. Hence, the continuity across the cut should not make much difference for how attention is deployed after the cut. If one regards priming of attention as a stimulus-driven principle (Theeuwes, 2013; but see Lamy & Kristjánsson, 2013), the fact that we failed to find an influence of visual similarity under passive viewing conditions (Experiment 3) could be regarded as evidence against a priming account. If the benefit of (visually similar) WSC only occur under active viewing conditions but not under more passive free-viewing conditions, then it is likely that the similarity-based continuity benefits are indeed mediated by the viewer's active role and a top-down controlled selection of visual content (cf. Fecteau, 2007). After all, even in highly controlled visual search experiments, priming effects are complemented by other memory processes that also rely on previous experience, including associative learning, working memory, episodic memory, or semantic memory (Chun & Jiang, 1998; Chun & Nakayama, 2000; Hutchinson & Turk-Browne, 2012; Olivers et al., 2006). Which of these different forms of memory contributed to the currently observed attention effects should be further investigated in future experiments.

## Conclusion

In the present study, we investigated how visual attention operates in edited dynamic scenes and how it interacts with memory to bridge the visual discontinuities after cuts. We found that visual similarity across cuts is decisive for how rapidly and effectively viewers are able to actively deploy their attention to a movie's continuation following a cut. The facilitative influence of visual similarity on attention takes effect at an early point in time within postcut images (Experiment 2), and supports the recognition of novel information in visually similar postcut images (Experiment 1). Given that cuts with a higher continuity in content, such as WSC, are often characterized by a higher degree of visual similarity compared to more discontinuous BSC, the beneficial effect of visual similarity on attention could also facilitate movie comprehension. This is an important finding given the abundance of edited visual material and the lack of systematic research in the rules underlying efficient allocation of attention across cuts.

*Keywords:* attention, continuity, dynamic scenes, editing, memory

## Acknowledgments

This research was supported by grant CS 11-009 from the Wiener Wissenschafts-, Forschungs- und

Technologiefonds (WWTF, Vienna Science and Technology Fund) awarded to Ulrich Ansorge, Shelley Buchinger, and Otmar Scherzer; and grant ERC-2010-AdG #269716 – MULTISENSE, awarded to Peter König. The authors thank the anonymous reviewers for their helpful and constructive feedback on previous versions of this manuscript. The authors further thank Stephanie Leeb and Markus Grüner for assistance with data collection, and Melanie Demir for assistance with stimulus preparation. Parts of this research were previously presented at the 14th and 15th Annual Meeting of the Vision Sciences Society (St. Pete Beach, FL, USA). The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Commercial relationships: none.

Corresponding author: Christian Valuch.

Email: christian.valuch@psych.uni-goettingen.de.

Address: Georg-Elias-Müller-Institute for Psychology, Georg-August-University Göttingen, Göttingen, Germany.

## References

- Açık, A., Bartel, A., & König, P. (2014). Real and implied motion at the center of gaze. *Journal of Vision*, 14(1):2, 1–19, doi:10.1167/14.1.2. [PubMed] [Article]
- Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision*, 11(8):9, 1–15, doi:10.1167/11.8.9. [PubMed] [Article]
- Ansorge, U., Buchinger, S., Valuch, S., Patrone, A. R., & Scherzer, O. (2014). Visual attention in edited dynamical images. *SIGMAP 2014: Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (pp. 198–205). SCITEPRESS Digital Library. doi:10.5220/0005101901980205
- Arijon, D. (1976). *Grammar of the film language*. Beverly Hills, CA: Simlan James Press.
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437–443, doi:10.1016/j.tics.2012.06.010. [PubMed]
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384, doi:10.3758/BF03192707. [PubMed]
- Berliner, T., & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video*, 63(1), 44–63, doi:10.1353/jfv.2011.0008.
- Böhme, M., Dorr, M., Krause, C., Martinetz, T., & Barth, E. (2006). Eye movement predictions on natural videos. *Neurocomputing*, 69(16), 1996–2004, doi:10.1016/j.neucom.2005.11.019.
- Bordwell, D., & Thompson, K. (2001). *Film art: An introduction* (6th ed.). New York: McGraw-Hill.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14325–14329, doi:10.1073/pnas.0803390105. [PubMed]
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436, doi:10.1163/156856897X00357. [PubMed]
- Carmi, R., & Itti, L. (2006a). The role of memory in guiding attention during natural vision. *Journal of Vision*, 6(9):4, 898–914, doi:10.1167/6.9.4 [PubMed] [Article]
- Carmi, R., & Itti, L. (2006b). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46(26), 4333–4345, doi:10.1016/j.visres.2006.08.019. [PubMed]
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71, doi:10.1006/cogp.1998.0681. [PubMed]
- Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, 7(1–3), 65–81, doi:10.1080/135062800394685.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159, doi:10.1037/0033-2909.112.1.155. [PubMed]
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617, doi:10.3758/BF03195489. [PubMed]
- Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1476–1490, doi:10.1037/a0027737. [PubMed]
- Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film.

- Psychological Science*, 21(3), 432–439, doi:10.1177/0956797610361679. [PubMed]
- Deubel H., & Schneider W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36, 1827–1837, doi: 10.1016/0042-6989(95)00294-4. [PubMed]
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10):28, 1–17, doi:10.1167/10.10.28. [PubMed] [Article]
- Dorr, M., Vig, E., & Barth, E. (2012). Eye movement prediction and variability on natural video data sets. *Visual Cognition*, 20(4–5), 495–514, doi:10.1080/13506285.2012.667456. [PubMed]
- D’Zmura, M. (1991). Color in visual search. *Vision Research*, 31(6), 951–966, doi:10.1016/0042-6989(91)90203-H. [PubMed]
- Fecteau, J. H. (2007). Priming of pop-out depends upon the current goals of observers. *Journal of Vision*, 7(6):1, 1–11, doi:10.1167/7.6.1. [PubMed] [Article]
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030–1044, doi:10.1037/0096-1523.18.4.1030. [PubMed]
- Frith, U., & Robson, J. E. (1975). Perceiving the language of films. *Perception*, 4(1), 97–103, doi:10.1068/p040097. [PubMed]
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, 10(13), 805–808, doi:10.1016/S0960-9822(00)00563-7. [PubMed]
- Geyer, T., Müller, H. J., & Krummenacher, J. (2008). Expectancies modulate attentional capture by salient color singletons. *Vision Research*, 48(11), 1315–1326, doi: 10.1016/j.visres.2008.02.006. [PubMed]
- Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, 37(7), 957–964, doi:10.1016/j.compbiomed.2006.08.018. [PubMed]
- Hasson, U., Landesman, O., Knappmeyer, B., Valines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections: The Journal of Movies and Mind*, 2(1), 1–26, doi:10.3167/proj.2008.020102.
- Helmholtz H. (1867). *Handbuch der physiologischen optik [Handbook of physiological optics]*. Leipzig, Germany: Voss.
- Hochberg, J., & Brooks, V. (1996). The perception of motion pictures. In Friedman M. P. & Carterette E. C. (Eds.), *Cognitive ecology* (2nd ed, pp. 205–292). London, UK: Academic Press.
- Horstmann, G. (2002). Evidence for attentional capture by a surprising color singleton in visual search. *Psychological Science*, 13(6), 499–505, doi:10.1111/1467-9280.00488. [PubMed]
- Horstmann, G. (2005). Attentional capture by an unannounced color singleton depends on expectation discrepancy. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1039–1060, doi:10.1037/0096-1523.31.5.1039. [PubMed]
- Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, 16(12), 576–579, doi:10.1016/j.tics.2012.10.003. [PubMed]
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306, doi:10.1016/j.visres.2008.09.007. [PubMed]
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897–1916. [PubMed]
- Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, 72(1), 5–18, doi:10.3758/APP.72.1.5. [PubMed]
- Lamy, D. F., & Kristjánsson, Á. (2013). Is goal-directed attentional guidance just intertrial priming? A review. *Journal of Vision*, 13(3):14, 1–19, doi: 10.1167/13.3.14. [PubMed] [Article]
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47(19), 2483–2498, doi:10.1016/j.visres.2007.06.015. [PubMed]
- Levin, D. T., & 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(4), 357–380, doi:10.1207/S1532785XMEP0204\_03.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490, doi: 10.3758/BF03210951. [PubMed]
- Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The

- tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS ONE*, 10(11), e0142474, doi:10.1371/journal.pone.0142474. [PubMed]
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110, doi:10.1023/B:VISI.0000029664.99615.94.
- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489–1517, doi:10.1111/j.1551-6709.2011.01202.x. [PubMed]
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, 22(6), 657–672, doi:10.3758/BF03209251. [PubMed]
- Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, 58(7), 977–991, doi:10.3758/BF03206826. [PubMed]
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1), 5–24, doi:10.1007/s12559-010-9074-z.
- Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top-down dimensional set. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 1–16, doi:10.1037/0096-1523.35.1.1. [PubMed]
- Münsterberg, H. (1916 /1996). *The photoplay—A psychological study*. New York: D. Appleton. Retrieved from: <http://www.gutenberg.org/files/15383/15383-h/15383-h.htm>
- Murch, W. (2001). *In the blink of an eye: A perspective on film editing* (2nd ed.). Beverly Hills, CA: Silman-James Press.
- Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243–1265, doi:10.1037/0096-1523.32.5.1243. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442, doi:10.1163/156856897X00366.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32A(1), 3–25, doi:10.1080/00335558008248231.
- R Core Team. (2015). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Reisz, K., & Millar, G. (1968). *The technique of film editing* (2nd ed.). London, UK: Focal Press.
- Ross, N. M., & Kowler, E. (2013). Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision*, 13(4):1, 1–19, doi:10.1167/13.4.1. [PubMed] [Article]
- Salt, B. (2009). *Film style and technology: History and analysis* (Vol. 3). Totton, UK: Starword.
- Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition*, 28(2), 214–223, doi:10.3758/BF03213801. [PubMed]
- Schwan, S., & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science*, 21(7), 970–976, doi:10.1177/0956797610372632. [PubMed]
- Shimamura, A. P., Cohn-Sheehy, B. I., & Shimamura, T. A. (2014). Perceiving movement across film edits: A psychocinematic analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 77–80, doi:10.1037/a0034595.
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, 6(1), 1–27, doi:10.3167/proj.2012.060102.
- Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, 2(2), 1–17, doi:10.16910/jemr.2.2.6.
- Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, 17(6–7), 1083–1108, doi:10.1080/13506280802678557.
- Smith, T. J., Levin, D., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science*, 21(2), 107–113, doi:10.1177/0963721412437407.
- Smith, T. J., & Martin-Portugues Santacreu, J. Y. (in press). Match-action: The role of motion and audio in creating global change blindness in film. *Media Psychology*. Advanced online publication, doi:10.1080/15213269.2016.1160789.
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal*

- of Vision, 13(8):16, 1–24, doi:10.1167/13.8.16. [PubMed] [Article]
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215, doi:10.1037/0278-7393.6.2.174. [PubMed]
- Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(1), 11–32, doi:10.1007/BF00130487.
- Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1), 1–20, doi:10.1016/S0010-0277(98)00026-2. [PubMed]
- Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 1–8, doi:10.1098/rstb.2013.0055. [PubMed]
- Thompson, K. (1999). *Storytelling in the new Hollywood*. Cambridge, MA: Harvard University Press.
- Valuch, C., & Ansorge, U. (2015). The influence of color during continuity cuts in edited movies: An eye-tracking study. *Multimedia Tools and Applications*, 74(22), 10161–10176, doi:10.1007/s11042-015-2806-z.
- Valuch, C., Ansorge, U., Buchinger, S., Patrone, A., & Scherzer, O. (2014). The effect of cinematic cuts on human attention. *TVX '14: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (pp. 119–122). New York: ACM. doi:10.1145/2602299.2602307
- Valuch, C., Becker, S. I., & Ansorge, U. (2013). Priming of fixations during recognition of natural scenes. *Journal of Vision*, 13(3):3, 1–22, doi:10.1167/13.3.3. [PubMed] [Article]
- Valuch, C., Seywerth, R., König, P., & Ansorge, U. (2015). “Why do cuts work?”: Implicit memory biases attention and gaze after cuts in edited movies. *Journal of Vision*, 15(12), 1237, doi:10.1167/15.12.1237. [Abstract]
- Vedaldi, A., & Fulkerson, B. (2010). VLFeat: An open and portable library of computer vision algorithms. *MM '10: Proceedings of the ACM International conference on Multimedia* (pp. 1469–1472). New York: ACM. doi:10.1145/1873951.1874249
- Võ, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12(13):3, 1–14, doi:10.1167/12.13.3. [PubMed] [Article]
- Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, 12(1):16, 1–27, doi:10.1167/12.1.16. [PubMed] [Article]
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Wolfe, J. M. (1994). Guided search 2.0—A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238, doi:10.3758/BF03200774. [PubMed]
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495–501, doi:10.1038/nrn1411. [PubMed]
- Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain’s cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, 4:168, doi:10.3389/fnhum.2010.00168. [PubMed]
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80–84, doi:10.1111/j.1467-8721.2007.00480.x. [PubMed]