

To appear in: *Art & the Senses*, ed. D. P. Melcher and F. Bacci. New York: Oxford University Press.

Film, Narrative, and Cognitive Neuroscience

Jeffrey M. Zacks & Joseph P. Magliano

## Introduction

Comprehending a film is an amazing feat of neural and cognitive processing. A series of still pictures are projected quickly on a screen, accompanied by a stream of sound—and a viewer has an experience that can be as engaging, emotionally affecting, and memorable as many experiences in real life. Comprehending film is all the more amazing when one considers how films are constructed. Films typically are comprised of hundreds and into the thousands of individual camera shot, which are continuous runs of the camera. Shot are edited together in such a way to create scenes, which are sequences of goal-directed actions or unintentional events that take place in a particular location (or one or two locations when events are depicted as occurring concurrently). Shots on average last only a few seconds and are edited together in such a way that the vast majority of them go unnoticed by the viewer (Magliano, Miller, & Zwaan, 2001; Bordwell & Thompson, 2003; Smith & Henderson, in press)

Consider Figure 1, which depicts about 15 seconds of film from the James Bond film, *Moonraker* (Gilbert, 1979). This shot sequence takes place after a fight in mid air between a villain, Jaws, and Bond, in which Bond escapes. Shot 1 depicts the villain pulling the ripcord for his parachute, which breaks. Shot 2 is a high angle long shot of the character reacting to the fact that the ripcord broke. Shot 3 depicts a reverse, high angle long shot and the character starts to flap his arms. Circus music, whose source is not yet apparent, starts during this shot and continues throughout the rest of the scene. Shot 4 is where things get particularly interesting from our perspective. This shot depicts a seemingly incongruent shot of the outside of a circus tent; Shot 5 returns to a medium shot of the character who continues to flap his arms. Shot 6 depicts the inside of a circus tent. At this point, many viewers generate the inference that Jaws will land on the circus tent (Magliano, Dijkstra, & Zwaan, 1996). It is impressive that despite the fact that the visual flow of information across shots in film (including those depicted in Figure 1) bears little resemblance to the perceptual flow of information as we interact in the real world (Cutting, 2005), viewers can process the juxtaposed and seemingly disparate camera shots and comprehend the story events depicted in them with ease. Viewers perceive spatial and temporal continuity across the shots in Figure 1 (Magliano et al., 2001), allowing them to infer that villain will fall into the circus tent. This example is particularly impressive because viewers must infer that the villain is spatially over the circus tent and that the events depicted in Shot 6 are actually taking place inside the location depicted in Shot 4.

Sequences such as this vividly illustrate that viewers construct representations of events in film that go far beyond the physical stimulus. How do the mind and brain transform a stream of flickering lights and oscillating speaker cones into a coherent world? Some parts of how people understand film can be explained by explaining how the components film are processed at perceptual, cognitive, and neurophysiological levels: Much of what is said about music and sound in Chapters 4.2 and 4.4, and about vision in Chapters 5.2, 5.3, and 5.6 carries over directly to figuring out how film works. The interaction of audition and vision, the focus of chapter 4.3, is also very important for film comprehension. But somehow audition, vision, and their interaction aren't whole story. This is because films do not simply present audiovisual signals. Rather, in most cases films are vicariously experienced events (Tan, 1995; Zillmann, 1995; Tan, 1996; Copeland, Magliano, & Radvansky, 2006; Magliano, Radvansky, & Copeland, 2007b). Events have their own psychology above and beyond the psychology of the sensory modalities. Thus, to understand how film works one needs to understand how experiences of events are constructed from auditory and visual signals. To do so, we will take a cognitive neuroscience approach, which means we will be interested in phenomena in terms of their information-processing properties and in terms of their neurobiological properties (for a different approach, see Chapter X). We will discuss both behavioral and neurological data that sheds light into how filmed events are processed and how comprehension emerges from those processes.

## Narration and Film

The technical details of auditory and visual reproduction in commercial cinema have evolved tremendously over the last hundred years, with visual frame rates going from 16 frames per second to 24, flicker rates tripling to 72 frames per second, visual contrast increasing, color being added, sound being added and going from one channel, to two, to five or six. This increased rate in the presentation of images does not change the essential task of a viewer; viewers must construct a mental representation of the depicted events. This situation is complicated in film given the fact that camera shots are often filmed at different times and places and that individual camera shots that contain minimal feature overlap can be edited together in such a way that viewers perceive continuity of action in spite of the lack of feature overlap (Bordwell, 1985; Bordwell & Thompson, 2003). (Shots 3 through 4 in Figure 1 illustrate this aspect of films.)

In order to tell a story in film, filmmakers rely on formal devices that separate it from activities “in the wild” (Cutting, 2005). Filmmakers adopt narrative devices, such as the cut, framing content in a shot, the placement of objects in the scene, directions to actors, music, sound, and dialog (Bordwell & Thompson, 2003). Perhaps the most significant of these is the cut, which is a term used to refer to the juncture between shots. At a cut, every point in the image changes discontinuously. How do human perceptual systems cope with such a jarring transition? According to classical film theory (Bordwell, 1985; Bordwell & Thompson, 2003), cuts can serve two distinct purposes. The majority of cuts are *continuity edits*, which serve to bridge incidental breaks in the physical features of activity as smoothly as possible in order to maintain continuity of action across spatiotemporal discontinuities. Techniques to do this include preserving the direction of motion across cuts, maintaining ongoing sounds in the soundtrack, and overlapping the objects and characters visible before and after the cut. Approximately 95%

of all cuts are continuity edits (Cutting, 2005). The remaining cuts are *scene breaks*, at which one action ends and another begins.

Continuity edits are valuable clues to how the brain processes events. This is particularly true for films made in the dominant “Hollywood style.” Hollywood style attempts to be invisible, so that the viewer is absorbed by the events of the story and does not much notice cuts, camera motion, camera angles, lighting, and so forth. When a filmmaker uses a technique (say, a cut) and we as viewers don’t notice, this is a hint that whatever is changed by the technique is not very salient to our perceptual and cognitive systems (Levin & Simons, 2000; Cutting, 2005).

Some authors have argued that cuts in general “work” because they correspond to visual interruptions that occur naturally due to the movements of the eyes, in particular *blinks* and *saccades* (Murch, 2001; Cutting, 2005). Blinks, of course, are the brief closures of the eye that typically occur several times a minute. Saccades are rapid ballistic movements of the eye that occur when we shift from looking at one thing to looking at another. Because both result in transient insensitivity to visual input, it is tempting to think they correspond to cuts. The film editor and director Walter Murch (2001, p. 62-63) proposed that “a shot presents us with an idea, or a sequence of ideas, and the cut is a ‘blink’ that separates and punctuates those ideas.” He went on to argue that if you were to watch the eyes of an audience viewing a well-edited film you would see them all blinking together at the cuts. The view that cuts correspond to the transient blindnesses caused by blinks or saccades is intuitive, but a quick investigation suggests it can’t be quite right. Try this: Watch on a film or TV program from a distance of at least six feet, so the image of the screen on you eyes is not too large. Fix your eyes on the middle of the screen and don’t move them. Don’t blink for 30 seconds or so. Did the cuts suddenly appear strange and jarring? We suspect not—and the limited experimental data available support this informal observation (Smith & Henderson, in press). So, although in natural viewing some cuts may be effectively hidden by saccades or blinks, this can’t be the whole story. Even without transient blindness, a visual change may be effectively camouflaged by another visual change that occurs near it in time. Called *masking*, this effect is pervasive and can lead to pronounced failures to detect changes (Rensink, O’Regan, & Clark, 1997; O’Regan, Rensink, & Clark, 1999). Many continuity edits may build in masking by placing the cut within or before a period of fast motion, as when a film editor *matches on action*. Match-on-action cuts have been found to be less detectable than cuts that aren’t followed by fast motion (Smith & Henderson, in press). So, cuts may be unobtrusive either because they correspond with brief blindnesses or because they are masked by other visual changes. We propose that these two mechanisms account for the unobtrusiveness of most continuity edits. But this still leaves a bunch of cuts left over—those that we seem not to notice despite large unmasked visual changes that don’t co-occur with blinks or saccades.

To account for how cuts function more broadly, we make a perhaps counterintuitive proposal: Some cuts may “work” (depending on how they are executed), because human perceptual systems are already segmenting ongoing activity into discrete events all the time. If a cut is placed where the observer would naturally segment the activity, then the cut will be experienced as natural even if it is readily detectable. Such cuts needn’t be hidden by an eye movement or visual masking—anything goes. To make this argument, we will have to explain a little bit about the psychology and neuroscience of event segmentation.

## Event Segmentation

There is good behavioral and neurophysiological evidence that when people watch ongoing activity, they segment it into meaningful events. Segmentation is an ongoing concomitant of normal perception—one that is related to eye movements and blinks, but that reflects the operation of a broader system for orienting attention and updating memory (for reviews, see Kurby & Zacks, 2008; Zacks & Swallow, 2007). To measure event segmentation behaviorally, one can ask a group of viewers to watch a film and press a button each time they feel one event has ended and another has begun (Newtson, 1973). Figure 2 gives an example of typical event boundary locations. Viewers agree well with each other on where the boundaries between events are located, and also show individual differences that are stable over time (Newtson & Engquist, 1976; Speer, Swallow, & Zacks, 2003). Viewers can segment events at various temporal grains; if the experimenter asks a viewer to identify more coarse-grained or fine-grained events, participants generally have no problem complying. Fine-grained events appear to be spontaneously grouped hierarchically into larger events (Zacks, Tversky, & Iyer, 2001b). In short, viewers' segmentation of films into events is reliable and regular, suggesting that the segmentation task taps into ongoing perceptual processing.

Neurophysiological evidence converges with these behavioral data to suggest that event segmentation happens on an ongoing basis during perception. The relevant studies share a common logic: Initially, viewers who are naïve to the event segmentation task watch films while some measure of neural activity is taken. Then, the viewers watch the films again and segment them into meaningful events. The event boundaries identified in this second phase are used as markers to interrogate the previously recorded brain activity recordings, allowing the researcher to ask what was changing in the neural activity during the initial viewing at those points that the viewer would later identify as event boundaries. An attractive feature of this method is that any neural changes observed at event boundaries are unlikely to reflect deliberate event segmentation, because during the initial viewing participants are unaware that they will later be asked to segment the activity. In several studies, this approach has been applied using fMRI, with stimuli including movies of everyday events (Zacks et al., 2001; Speer et al., 2003), a French art film (Zacks, Swallow, Speer, & Maley, 2006a), and simple animations (Zacks, Swallow, Vettel, & McAvoy, 2006c). In all cases the fMRI data showed transient increases at event boundaries in activity in a distributed network (Figure 3). Similar results have also been observed using electroencephalography (EEG), which provides a measure of the large-scale electrical activity of the brain (Sharp, Lee, & Donaldson, 2007). Finally, researchers have used pupil diameter as a measure of cognitive load during film viewing (Swallow & Zacks, 2004). In studies of memory and problem-solving, pupil diameter has been found to increase as the cognitive requirements of a task increase (Beatty & Lucero-Wagoner, 2000). In this study, transient increases in pupil diameter were observed around those times that viewers would later identify as event boundaries. Thus, these neurophysiological data converge with the behavioral data to suggest that viewers spontaneously segment ongoing activity into events as a normal concomitant of perception.

What determines when observers perceive event boundaries to occur? One recent theory (Zacks, Speer, Swallow, Braver, & Reynolds, 2007) proposes that, as part of ongoing understanding, observers make predictions about what will happen next in an activity. A particular individual

doesn't make predictions about everything that might happen in the activity, everything that might change. Rather, viewers monitor a set of features of activity that are salient and relevant to their goals. When predictions about these features are violated, viewers perceive the onset of a new event.

According to this theory, event boundaries should tend to occur when features in an activity change, because in general changes are less predictable than stasis (though there are some exceptions).

What sorts of changes matter? Viewers tend to segment activity at physical changes—changes in the movement of actors and objects, in spatial location, and in time (Newtson, Engquist, & Bois, 1977; Magliano et al., 2001; Zacks, 2004; Hard, Tversky, & Lang, 2006). For example, event boundaries tend to occur when objects or body parts are accelerating relative to each other and moving fast (Zacks, 2004; Zacks, Kumar, & Abrams, under review). If movement is important for event segmentation, one would expect that activity in brain areas specialized for movement processing would be related to event segmentation. This appears to be the case. In particular, area MT+ is an area in the lateral posterior cortex that responds selectively to motion (Tootell et al., 1995; Chawla, Phillips, Buechel, Edwards, & Friston, 1998). As can be seen in Figure 3, MT+ is strongly activated during event boundaries (Speer et al., 2003) such that there is a three-way relationship between movement, event segmentation, and activity in MT+: When things move quickly, activity in MT+ is greater and people tend to perceive event boundaries (Zacks et al., 2006c).

In addition to physical changes, viewers tend to segment activity at conceptual changes, such as character goals and causal relationships. These have also been found to predict where event boundaries will occur (Magliano, Taylor, & Kim, 2005; Zacks, Speer, & Reynolds, in press). The neural processes that relate conceptual changes to event segmentation are less well understood than those relating movement to event segmentation. However, we do know that changes to both physical and conceptual features predict activity in most of areas that increase at event boundaries (Zacks et al., 2006a; Speer, Reynolds, & Zacks, 2007; Speer, Reynolds, Swallow, & Zacks, under review). This supports the idea that we perceive event boundaries because we process changes in the features to which we are attending.

## Cuts and Continuity

With this account of event segmentation ready to hand are in a position to return to the question of how cuts function. We propose that continuity edits and cuts at scene breaks follow different rules. For continuity edits the cut must be hidden by a blink, saccade, visual masking or something else in order to “work.” For scene breaks, however, it is not necessary to visually hide the cut if it happens at an event boundary. This proposal makes a clear proposal: In a well-edited movie, cuts that classical film theory would identify as scene breaks should be identified as event boundaries, whereas continuity edits should not.

We recently set out to test this using behavioral and fMRI data collected while viewers watched *The Red Balloon* (Lamorisse, 1956), a French art film about a boy who befriends the balloon of the title. *The Red Balloon* is 33 min long and contains 214 cuts. Each cut was categorized based

on whether it introduces a change in spatial location (e.g., moving from indoors to outdoors or from a trolley to the street), a change in time (usually jumping forward to elide an unimportant part of the activity), or a change in the action being performed. The remainder of cuts produced changes in viewpoint on the scene but did not change the temporal or spatial location. According to classical film theory, a scene break occurs when a new action is performed; scene breaks often co-occur with spatial or temporal changes, but this is not necessary.

Viewers segmented the film into coarse and fine events. We used the locations and types of cuts to ask how editing was related to behavioral segmentation and to brain activity during film viewing. The event segmentation data supported classical film theory's account of continuity editing. For both coarse and fine segmentation, cuts that introduced action discontinuities were quite likely to be perceived as event boundaries. Cuts that introduced spatial or temporal discontinuities, but preserved continuity of action, were associated with fine-grained event boundaries but not with coarse-grained boundaries. One possibility is that fine-grained segmentation depends more on the processing of physical changes, whereas coarse-grained segmentation is more sensitive to conceptual changes. Notably, cuts in and of themselves were not associated with event boundaries—if a cut merely changed the camera viewpoint within a scene, it had no discernable effect on coarse-grained segmentation and only a small effect on fine-grained segmentation. In contrast, viewers perceived event boundaries in both the fine and coarse segmentation task if the cuts coincided with a change in action. These data suggest that—at least for *The Red Balloon*—continuity editing techniques are successful in perceptually smoothing over full-field visual discontinuities and that scene boundaries require a break in action.

How does the brain achieve this perceptual smoothing over? The fMRI data provide some hints. We analyzed the fMRI response to cuts as a function of whether the cut was perceived as an event boundary and whether it introduced a spatial or temporal change. One set of brain regions selectively increased in activity at those cuts that were event boundaries (shown in blue in figure 4). These regions overlapped with those that had previously been found to respond at event boundaries, including regions in the lateral temporal and parietal cortex. A different set of regions selectively increased in activity at those cuts that were not judged to be event boundaries but that did have spatial or temporal changes (shown red in figure 4). These regions included the mid-cingulate gyrus, a region in the lateral inferior parietal lobule, and the lateral anterior temporal lobes. We cannot be sure exactly how these regions participate in bridging spatiotemporal discontinuities in these types of cuts based just on these data, but these results do suggest further lines of experimentation. The activated regions overlap with a network that has been identified with attentional control with tasks, in distinction to switches from task to task (Dosenbach et al., 2007). One possibility is that bridging continuity gaps in movies requires the same processes as within-task attentional control. This fits with the proposal that continuity editing works by hiding cuts using blinks, saccades, and masking—all three are associated with attentional reorienting. This hypothesis could be tested using task batteries designed to assess different aspects of attentional control directly (e.g., Fan, McCandliss, Sommer, Raz, & Posner, 2002)

The studies just described were correlational—we did not experimentally manipulate the locations or types of cuts, but took them as they came from the director. Experimental studies

also suggest that cuts are less detectible when they correspond to event boundaries. In a study of infant perception (Baldwin, Baird, Saylor, & Clark, 2001), infants watched a short film of a woman cleaning a kitchen until they become bored and looked away. They then were tested with versions of the film that had been modified by introducing brief pauses, placed either to correspond with event boundaries or slightly before or after. The infants looked longer at the versions with the pauses at nonboundaries, suggesting that placing pauses before or after the event boundary made the films more different to them than placing the pauses at event boundaries. Another experiment asked adults to detect cuts in movies presented either intact or with the order of shots scrambled (d'Ydewalle & Vanderbeeken, 1990). Scrambling shot order disrupts the structure of events, and thus we would expect it to make cuts more detectible. This was what was found. In a study of memory for events depicted in film, viewers watched brief film clips that contained zero or one cuts, and consisted of one or two actions (Carroll & Bever, 1976). After each clip they were shown a six-frame excerpt from the sequence they had just seen or from a similar clip, and asked to distinguish whether the excerpt came from the clip they had watched. The excerpts that did come from the just-watched clip were selected from either the first or second half. Excerpts from the first half were recognized more slowly than those from the second half, particularly when the second half included a new action. This suggests that when a new action had begun viewers perceived an event boundary, consistent with the segmentation results. Also consistent with the segmentation results, cuts themselves had little effect on recognition. (For reasons that are not clear, however, the authors interpret this as consistent with the idea that cuts themselves produced psychological boundaries.) In a final study (Schwan, Garsoffky, & Hesse, 2000), viewers watched movies of everyday events with cuts placed either at event boundaries or at points in between the boundaries (based on previously collected normative data). Consistent with the correlational studies, cuts in and of themselves had little effect on where viewers identified event boundaries. Cuts also had relatively little effect on memory: They accentuated memory for details at the locations of the cuts themselves, but did not affect memory for other points in the films.

## **Beyond Cuts—Commercials and Pauses**

Film theorists sometimes speak of cuts—at least in mainstream narrative film—as being largely invisible. Indeed, a major goal of continuity editing is to render cuts unobtrusive. However, film and television make use of a number of more intrusive transitions as well. These include fades, wipes, and iris effects. More rarely, a film can include pauses in which the screen is blank for a brief interval. In commercial television, segments of activity in a program are interrupted by commercial breaks. If cuts themselves don't affect event segmentation much, perhaps these more intrusive transitions do? A small number of experiments suggest this is the case. In a pair of studies conducted by Boltz (Boltz, 1992; Boltz, 1995), viewers watched an episode of a detective drama interrupted by zero, three, or six commercial breaks. Breaks were placed either so as to occur at natural event boundaries, or at the points in between event boundaries. The dependent measures included retrospective judgments of the length of the drama (Boltz, 1995) and tests of what had happened in the film and when (Boltz, 1992). Commercial breaks at event boundaries improved performance on both measures—the more commercials the better viewers' memory. Conversely, commercial breaks at non-boundaries hurt performance.

Schwan and colleagues applied a similar logic in a study using briefer films of everyday events (Schwan et al., 2000). Films were edited to elide sections of the action, either at event boundaries or at the middles between event boundaries. The deleted intervals were replaced with 1-second pauses. After viewing each movie, participants attempted to recall as much as possible. Deleting intervals at event boundaries had little effect on recall. However, deleting intervals in between event boundaries reduced memory for the remaining sections of the movie, and disrupted the temporal organization of memory.

These studies suggest that more intrusive editing techniques may affect event segmentation and memory. However, no studies to date have directly investigated common editing techniques such as fades and wipes. Commercial breaks are more extended, and do not come into play in current popular or art cinema. Pauses (blank screens) are closer to commonly used editing techniques, but themselves are rarely used. In the future it would be valuable to directly investigate the effects of intrusive editing techniques as they are actually used in cinema on event segmentation and memory.

In sum, some cuts in film may “work” by being camouflaged in blinks, saccades, or other visual transients, but other cuts may work because they co-opt an ongoing process of segmenting continuous activity into discrete events. As an artist, the director or editor has leverage to work with the natural tendencies of the perceptual system or against those tendencies. Working with natural tendencies will tend to produce smooth, naturalistic continuity editing. Working against them may introduce jarring perceptual effects, comprehension gaps, and memory encoding difficulties. Poorly used, these may frustrate the viewer, but when used to deliberate effect they may afford a richer and stranger cinematic experience. For example, a *jump cut* is a continuity edit that is not well masked and in which an object or person appears in two nearby screen locations on either side of the cut (Anderson, 1996). It produces the sensation that something has “jumped.” In *Dancer and the Dark* (von Trier.), Lars von Trier uses a hand-held camera to film the just-jailed protagonist exploring her cell, and several times drops a few frames to produce a jump cut. These jumps disorient the viewer slightly and deepen the sense that time passes very slowly in the cell.

## Narrative Events

When a viewer segments activity—in everyday life, or in a film—into discrete events, what sort of representation results? Here we draw on theories of narrative comprehension that have been developed and tested in the context of reading and listening as a basis for understanding film comprehension (van Dijk & Kintsch, 1983; Kintsch, 1988; Zwaan & Radvansky, 1998).

Although language processing clearly differs in important ways from film processing, it has proven useful to make the working assumption that both text processing and film understanding lead to similar representations of narrated events (Carroll, 1980; Bordwell, 1985; Branigan, 1992; Magliano et al., 2001; Copeland et al., 2006; Magliano, Zacks, Swallow, & Speer, 2007).

Reading comprehension—and by hypothesis film comprehension—involves constructing a multilevel representation (Kintsch & van Dijk, 1978). Readers construct a representation of the *surface structure* of a text, which captures the particular words and phrasing, but this representation is fleeting. In film, this would be analogous to a representation that included visual information about the details of characters’ clothing and props. Readers also construct a

representation of explicit text content, which is referred to as the *textbase*. This representation contains a network of propositions that is incrementally constructed during reading and represents the explicit ideas contained in a text). Film viewers likely construct a similar representation as well (Baggett, 1979). For example, a representation of the events depicted in this event sequence may consist of several propositions, such as (pull: ripcord, Jaws), (fall: ripcord), (flap: arms, Jaws). Relationships between propositions in this aspect of the representation can be established through co-reference. So, in the sequence shown in Figure 1, continuity can be established because Jaws is present in shots 1, 2, 3, and 5. However, deep meaning emerges with the construction of a model of the situation described by the text (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998; Magliano, Zwaan, & Graesser, 1999). This is usually referred to as a *situation model* (sometimes “mental model”). Here we will use the more general term *event model* (Zacks et al., 2007) to emphasize that we mean an event representation that could be derived from text, from a film, or from real life.

An event model provides an index of how story events are related along a number of dimensions, such as agents and objects, temporality, spatiality, causality, and intentionality (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998; Magliano et al., 1999; Zacks et al., in press). For example, comprehending the events depicted in Figure 1 requires viewers to generate spatial temporal relationships between Jaws falling without a parachute and the events depicted in the circus tent. These inferences afford a predictive inference that Jaws will land on the circus tent. These inferences are conceptualized as being part of the event model because they go beyond the propositions that are derived to reflect the explicitly seen events. Finally, narrative plots typically consist of a hierarchy of episodes that are causal related with one another (Thorndyke, 1977; Trabasso, Van den Broek, & Suh, 1989), and viewers must be able to construct representations that reflect the implicit plot structure. Just as is the case with narrative texts (e.g., Trabasso et al., 1989), viewers must infer and represent the causal relationships between events within a scene and across scenes, the latter of which provide a basis for understanding the plot structure). For example, in a stereotypical action-adventure such as a James Bond film, the character has a primary goal to stop the villain, which can only be accomplished by achieving a series of subordinate goals. If viewers cannot infer that Bond initially visits the villain in order to find out information regarding his involvement in the mystery he is to solve, then they will not be able to comprehend the film.

Neuroimaging studies indicate that narrative comprehension during reading relies on a distributed text-processing network, and also on neural systems that are selectively activated when processing texts that allow one to construct a coherent event model (Ferstl, 2007; Ferstl, Neumann, Bogler, & von, 2008). One important consequence of processing meaningful coherent text is that brain changes throughout the language network increase in strength (Yarkoni, Speer, & Zacks, 2008). However, there also is evidence that weaving a set of sentences into a coherent discourse selectively activates an area in the prefrontal cortex, specifically on the medial surface near the front of the anterior portion (Ferstl et al., 2008; Yarkoni et al., 2008). This area is usually referred to as dorsomedial prefrontal cortex, or dmPFC. One possibility is that dmPFC is selectively involved in the inference processes that allow one to fill out an event model based on a propositional textbase. Data consistent with this hypothesis come from a study that measured the response to changes in agents, objects, space, time and goals during narrative reading (Speer

et al., under review). The dmPFC was selectively activated when these situational dimensions changed. Interestingly, a portion of this region increased in activity as more dimensions changed, but did not show a strong increase for any particular change. Other regions of the brain responded selectively when particular dimensions of the situation changed; we will return to these in the following section.

In sum, event models for narratives reflect the events that are explicit and implicit parts of the narrative, who are involved in them, what their goals and what to do to achieve those goal, and the outcome of those activities within the spatial temporal framework of the story world (Zwaan & Radvansky, 1998; Magliano et al., 1999). Constructing these representations may depend on the dmPFC as well as other brain regions. This sort of conceptual information is relatively abstract—it seems on its face removed from the perceptual experience of watching or reading and the motor experience of acting in the world. How do the contents of event models relate to perceptual and motor experiences? We turn next to this question.

## Embodied Event Representations

Traditionally in artificial intelligence and cognitive psychology, event models were thought to be something like a description in a formal logic or a computer program, specifying parts of the event and features in an abstract language (e.g., Minsky, 1972). More recently, however, cognitive scientists and cognitive neuroscientists have come to view event representations as simulations of the situations they represent, which preserve some of the perceptual and motor details of the activity in a form that is closer to perception than to logic (Barsalou, Simmons, Barbey, & Wilson, 2003). This view is referred to as the *embodied cognition* or *perceptual symbol* view. In this section we will briefly review evidence for the perceptual and motor properties of event models and discuss how these are important for understanding film.

There have been a number of demonstrations that perceptual and motor features are activated automatically during reading. For example, in one study (Zwaan, Stanfield, & Yaxley, 2002), participants read sentences such as “The ranger saw an eagle in the sky” or “The ranger saw an eagle in its nest,” and then verified whether a line drawing matched one of the words in the sentence. After reading the sentence about the eagle in the sky, participants were relatively faster to verify a picture of an eagle if it were depicted with its wings outstretched. However, after reading the sentence about the eagle in the nest, participants were relatively faster to verify a picture of an eagle shown sitting with its wings folded. Such data suggest that during reading participants constructed representations that included perceptual information about the spatial configuration of the scene.

Neurophysiological data also support the construction of perceptual-motor content during reading. Some action words such as “lick,” “pick,” and “kick” are strongly associated with movements of particular parts of the body—in this case the tongue, hand, and leg, respectively. Reading such words selectively activates those parts of somatosensory and motor cortex that are activated when participants actually move their tongues, hands and legs (Hauk, Johnsrude, & Pulvermuller, 2004). Object concepts as well as verb concepts are represented in terms of their perceptual and motor properties (Martin, 2007). Similar regions in the temporal cortex are activated by pictures of objects and by their names, and (as noted previously) different categories

of objects are associated with activity in different regions in temporal cortex. Concepts of objects and actions are interrelated, such that the regions activated when thinking about tools are more strongly associated with regions activated during tool use than are the regions activated when thinking about non-manipulable objects (Johnson-Frey, 2004). Most of the existing data come from studies in which participants view a simple word, phrase, or picture, and make an explicit judgment about the stimulus. However, recent studies have found evidence that perceptual and motor contents are activated during ongoing story reading (Zacks et al., in press) and film viewing (Zacks et al., 2006a).

According to this view, whether one experiences an activity in real life, in a film, or through a book, one creates a set of conceptual representations that include perceptual and motor content. If this is true, and if people segment activity into events when features of the activity are changing unpredictably (as we argued above; see Event Segmentation), then readers should break up stories into events at changes just as film viewers do. In particular, the account of event models given in the previous section proposes that readers and viewers both segment activity into events when there are changes in the dimensions of the situation represented in their event models—agents and objects, temporality, spatiality, causality, and intentionality. The data seem to support this proposal: Whether segmentation is studied in story reading or film viewing, and whether it is measured directly or through indirect measures such as reading time, people appear to segment activity when features in the situation are changing (Zwaan, Radvansky, Hilliard, & Curiel, 1998; Magliano et al., 2001; Rinck & Weber, 2003; Speer & Zacks, 2005; Magliano et al., 2005; Zacks et al., in press). In neuroimaging studies, selective increases in brain activity are observed at these points, both for reading (Speer et al., 2007; Speer et al., under review) and for film viewing (Zacks et al., 2006a).

In short, when people understand activity they appear to construct event models that represent what is happening at any given moment. Event models include perceptual and motor content, which may be constructed from language or by inference, as well as directly experienced. Event models are updated at changes in features in the activity, a finding that brings together theories of narrative comprehension with theories of event segmentation.

## Conclusions

What does all of this say for understanding the film viewing experience? First, film directors and editors in the Hollywood Style exploit properties of perception and attention to create unobtrusive cuts. They can do this, we propose, in two distinct ways. One way is to hide a cut using visual masking or by diverting attention. Another way is to make a cut coincide with a perceptual event boundary.

A second conclusion is that perceptual event boundaries mark the major units of narrative comprehension, in film as well as in other media. The nervous system appears to devote a lot of effort to assembling representations of coherent events. Event representations are individuated based on features on multiple dimensions. Previous research has focused on agents, objects, space, time, causes, and goals. We believe—though we can present no evidence for it at this point—that these dimensions are instances of a broader principle: Event representations are individuated based on whatever features are important to the viewer's task. Agents, objects, and so forth are often found to be associated with comprehension because these dimensions often are

important for understanding. When structural features of film align with the situational features that define events, understanding is easier and memory is better. We believe that thinking about the alignment between structural features in film and situational features in events may allow new creative insights for film artists. By aligning structural elements with situational features one can create a narrative that is comprehensible and flows smoothly. By misaligning the two, one may deliberately disorient the viewer. In this vein, it is interesting to think about commercials in broadcast television such as situation comedies and serial dramas. Our hunch is that commercials often are placed just after situational changes that would typically be perceived as event boundaries. This may encourage viewers to remain in their seats through commercial breaks, but also may reduce comprehension.

Although the psychological and neurophysiological findings we have reviewed here can greatly inform our understanding of the experience of film, they are also greatly incomplete. For one thing, we have said nothing about viewers' affective response to film—though some would argue this is what film (and all forms of narration for that matter) is all about (Brewer & Lichtenstein, 1982; Oatley, 1995; Tan, 1995; Zillmann, 1995; Tan, 1996). We have done this reluctantly to restrain the scope of the chapter—there is much that could be said. Movies have been widely used to study the emotions in perception, cognition, and memory, because they are effective inducers of emotional responses (e.g., Tan, 1996). One important bridge between the perceptual-cognitive aspects of film understanding we have discussed here and emotion is the response of surprise, which has both cognitive and affective components (Brewer & Lichtenstein, 1982; Brewer & Ohtsuka, 1988). For a thoughtful exploration cognitive neuroscience of emotion as it applies to film, we recommend a recent volume by Greg Smith (2003).

The results reviewed here are also incomplete because our knowledge is incomplete. As those data reveal, the cognitive neuroscience of perception and cognition has much to tell us about the film viewing experience because our experience of film shares much with our experience of real life. However, some aspects of film are unique, and about these aspects our psychological and neurophysiological knowledge is much less well fleshed out. Here is a partial list of technical features of film that are simply begging for greater psychological and neurophysiological understanding: rate of cutting (Hochberg & Brooks, 2006), viewing angle (Cutting, 1987), camera motion, jump cuts, temporal reordering, relations between music and narrative action, deviations from perfect audiovisual synchrony, dubbing and subtitles. Clearly, the cognitive neuroscience of film is just in its infancy.

## References

- Anderson, J. (1996). *The reality of illusion: an ecological approach to cognitive film theory*. Carbondale: Southern Illinois University Press.
- Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning & Verbal Behavior*, 18(3), 333-356.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, 72(3), 708-717.
- Barsalou, L. W., Simmons, K. W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, 7, 84-91.
- Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 142-162). New York: Cambridge University Press.
- Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, 57(7), 1080-1096.
- Boltz, M. (1992). Temporal accent structure and the remembering of filmed narratives. *Journal of Experimental Psychology: Human Perception & Performance*, 18(1), 90-105.
- Bordwell, D. (1985). *Narration in the fiction film*. University of Wisconsin Press.
- Bordwell, D. & Thompson, K. (2003). *Film Art: An Introduction*. New York: McGraw-Hill.
- Branigan, E. (1992). *Narrative comprehension and film*. London ; New York: Routledge.
- Brewer, W. F. & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, 6(5-6), 437-486.
- Brewer, W. F. & Ohtsuka, K. (1988). Story structure and reader affect in American and Hungarian short stories. In C. Martindale (Ed.), *Psychological approaches to the study of literary narratives* (pp. 133-158). Hamburg: Helmut Buske Verlag.
- Carroll, J. M. & Bever, T. G. (1976). Segmentation in cinema perception. *Science*, 191(4231), 1053-1055.
- Carroll, J. M. (1980). *Toward a structural psychology of cinema*. The Hague ; New York: Mouton Publishers.
- Chawla, D., Phillips, J., Buechel, C., Edwards, R., & Friston, K. J. (1998). Speed-dependent motion-sensitive responses in V5: An fMRI study. *NeuroImage*, 7(2), 86-96.
- Copeland, D. E., Magliano, J. P., & Radvansky, G. A. (2006). Situation models in comprehension, memory and augmented cognition. In C. Forsythe, M. L. Bernard, & T. E. Goldsmith (Eds.), *Cognitive systems: Human cognitive models in systems design* (pp. 37-66). Mahwah, NJ: Erlbaum.
- Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception & Performance*, 13(3), 323-334.
- Cutting, J. E. (2005). Perceiving scenes in film and in the world. In J. D. Anderson & B. F. Anderson (Eds.), *Moving image theory: Ecological considerations* (pp. 9-27). Carbondale, IL: Southern Illinois University Press.
- d'Ydewalle, G. & Vanderbeeken, M. (1990). Perceptual and cognitive processing of editing rules in Ffilm. In R. Groner, G. d'Ydewalle, & R. Parham (Eds.), *From eye to mind: Information acquisition in perception, search, and reading* (pp. 129-139).
- Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. et al. (2007). Distinct brain networks for adaptive and stable task control in humans.

- Proceedings of the National Academy of Science of the United States*, 104(26), 11073-11078.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14(3), 340-347.
- Ferstl, E. C. (2007). The functional neuroanatomy of text comprehension: What's the story so far? In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher Level Language Processes in the Brain: Inference and Comprehension Processes* (pp. 53-102). Mahwah, NJ: Erlbaum.
- Ferstl, E. C., Neumann, J., Bogler, C., & von, C., DY. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29(5), 581-593.
- Gilbert, . L., (Director). (1979). *Moonraker* [Motion picture] United Kingdom: Danjaq.
- Hard, B. M., Tversky, B., & Lang, D. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, 34(6), 1221-1235.
- Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301-307.
- Hochberg, J. & Brooks, V. (2006). Film cutting and visual momentum. In M. A. Peterson, B. Gillam, & H. A. Sedgwick (pp. 206-228). New York: Oxford University Press.
- Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, 8(2), 71-78.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kurby, C. A. & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72-79.
- Lamorisse, A., (Writer/Director). (1956). *The red balloon* [Motion picture] France: Films Montsouris.
- 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.
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15(5), 533-545.
- Magliano, J. P., Zacks, J. M., Swallow, K. M., & Speer, N. K. (2007). Continuity editing in narrative film and the perception of event boundaries. *Abstracts of the Psychonomic Society*, 12, 27.
- Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1999). The role of situational continuity in narrative understanding. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 219-245). Mahwah, N. J.: Lawrence Erlbaum Associates.
- Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Predictive inferences in movies. *Discourse Processes*, 22, 199-224.
- Magliano, J. P., Radvansky, G. A., & Copeland, D. E. (2007b). Beyond language comprehension: Situation models as a form of autobiographical memory. *Higher level language processes in the brain: Inference and comprehension processes*.

- Magliano, J. P., Taylor, H. A., & Kim, H. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition*, 33(8), 1357-1367.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58(1), 25-45.
- Minsky, M. (1972). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.
- Murch, W. (2001). *In the blink of an eye : a perspective on film editing* (2nd ed.). Los Angeles: Silman-James Press.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28-38.
- Newton, D. & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12, 436-450.
- Newton, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35(12), 847-862.
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of "mudsplashes." *Nature*, 398(6722), 34.
- Oatley, K. (1995). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23(1-2), 53-74.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373.
- Rinck, M. & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory and Cognition*, 31(8), 1284-1292.
- 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.
- Sharp, R. M., Lee, J., & Donaldson, D. I. (2007). *Electrophysiological correlates of event segmentation: How does the human mind process ongoing activity?* Paper presented at the Annual Meeting of the Cognitive Neuroscience Society, New York.
- Smith, G. M. (2003). *Film structure and the emotion system*. Cambridge, UK ; New York: Cambridge University Press.
- Smith, T. J. & Henderson, J. M. (in press). Edit blindness: The relationship between attention and global change in dynamic scenes. *Journal of Eye Movement Research*.
- Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (under review). Reading stories activates neural representations of perceptual and motor experiences.
- Speer, N. K., Reynolds, J. R., & Zacks, J. M. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, 18(5), 449-455.
- Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective & Behavioral Neuroscience*, 3(4), 335-345.
- Speer, N. K. & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, 53, 125-140.
- Swallow, K. M. & Zacks, J. M. (2004). Hierarchical grouping of events revealed by eye movements. *Abstracts of the Psychonomic Society*, 11, 81.
- Tan, E. S. (1995). Film-induced affect as a witness emotion. *Poetics*, 23(1-2), 7-32.
- Tan, E. S. (1996). *Emotion and the structure of narrative film*. Lawrence Erlbaum Associates Hillsdale, NJ.

- Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9(1), 77-110.
- Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J. et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal Of Neuroscience*, 15(4), 3215-3230.
- Trabasso, T., Van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12(1), 1-25.
- van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- von Trier, L., (Writer/Director). *Dancer in the dark Denmark*: Zentropa Entertainments.
- Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *Neuroimage*, 41, 1408-1425.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979-1008.
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M. et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), 651-655.
- Zacks, J. M., Kumar, S., & Abrams, R. A. (under review). Using movement and intentions to understand human activity.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (in press). Situation changes predict the perception of event boundaries, reading time, and perceived predictability in narrative comprehension. *Journal of Experimental Psychology: General*.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, 133(2), 273-293.
- Zacks, J. M. & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80-84(5).
- Zacks, J. M., Swallow, K. M., Speer, N. K., & Maley, C. J. (2006a). The human brain's response to change in cinema. *Abstracts of the Psychonomic Society*, 11, 9.
- Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006c). Visual movement and the neural correlates of event perception. *Brain Research*, 1076(1), 150-162.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001b). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130(1), 29-58.
- Zillmann, D. (1995). Mechanisms of emotional involvement with drama. *Poetics*, 23(1-2), 33-51.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6(5), 292-297.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(2), 386-397.
- Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185.
- Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, 2(3), 199-220.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13(2), 168-171.

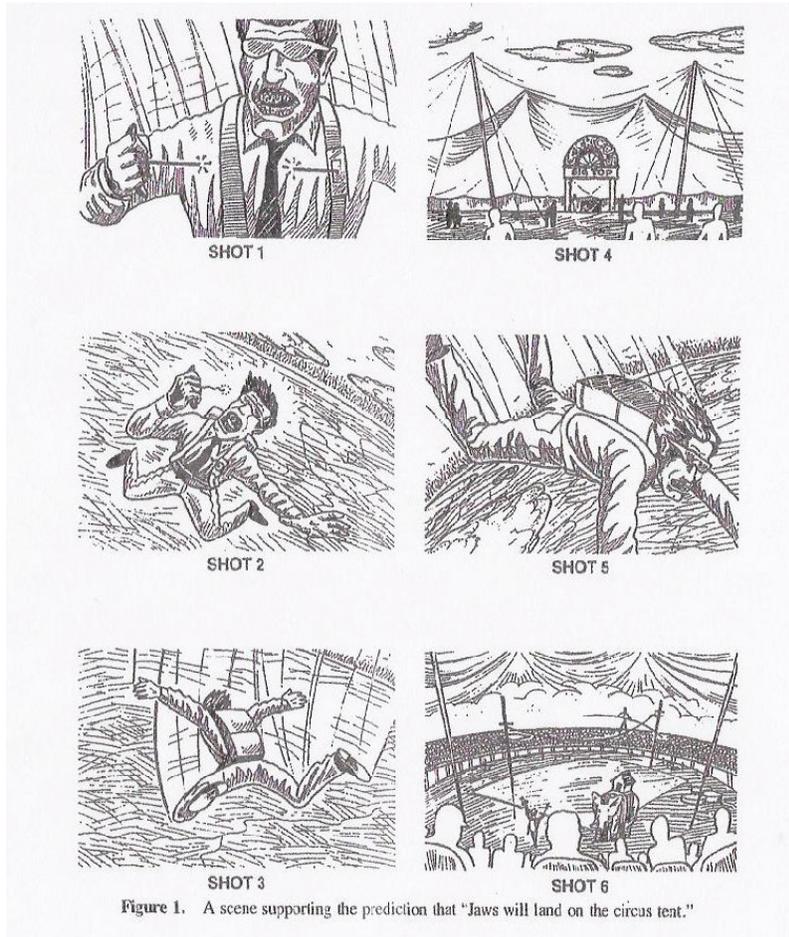
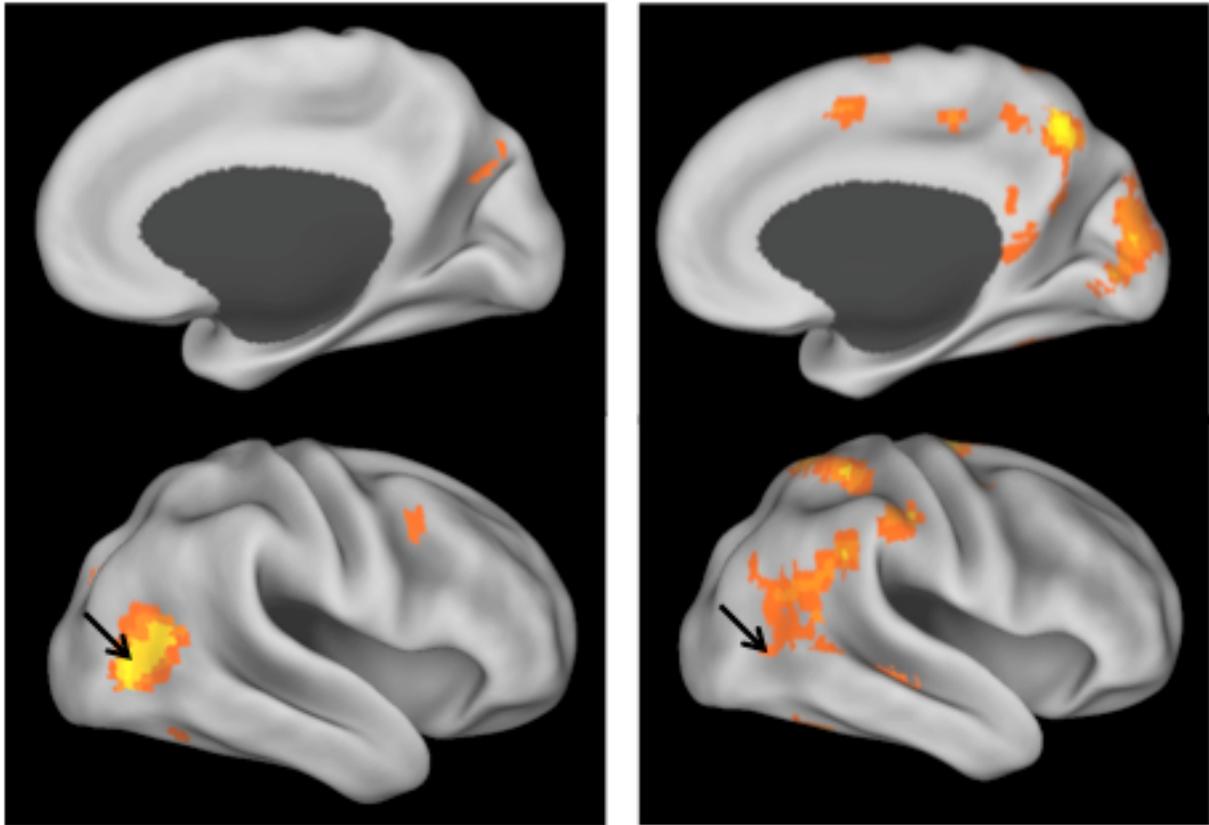


Figure 1. A story board depicting a scene from the James Bond film, *Moonraker*.



Figure 2. Example of event boundaries. These frames show the six coarse-grained event boundaries selected most frequently by a group of younger and older adults (Zacks, Speer, Vettel, & Jacoby, 2006, exp. 2). These boundaries marked the ends of events that could be described as: 1) Put down the tent. 2) Spread it out. 3) Insert the front tent pole. 4) Stake out the ends of the tent. 5) Stake out the sides. 6) Attach the rain fly. **(Reproduced with permission from Zacks & Swallow, 2007)**



Everyday Events

French Cinema

Figure 3. Evoked responses at event boundaries in laboratory-made movies of everyday events, and in a narrative French art film (Lamorisse, 1956). Data from (Zacks et al., 2001, left) and (Zacks et al., 2006c, right). Arrows indicate the approximate location of area MT+.

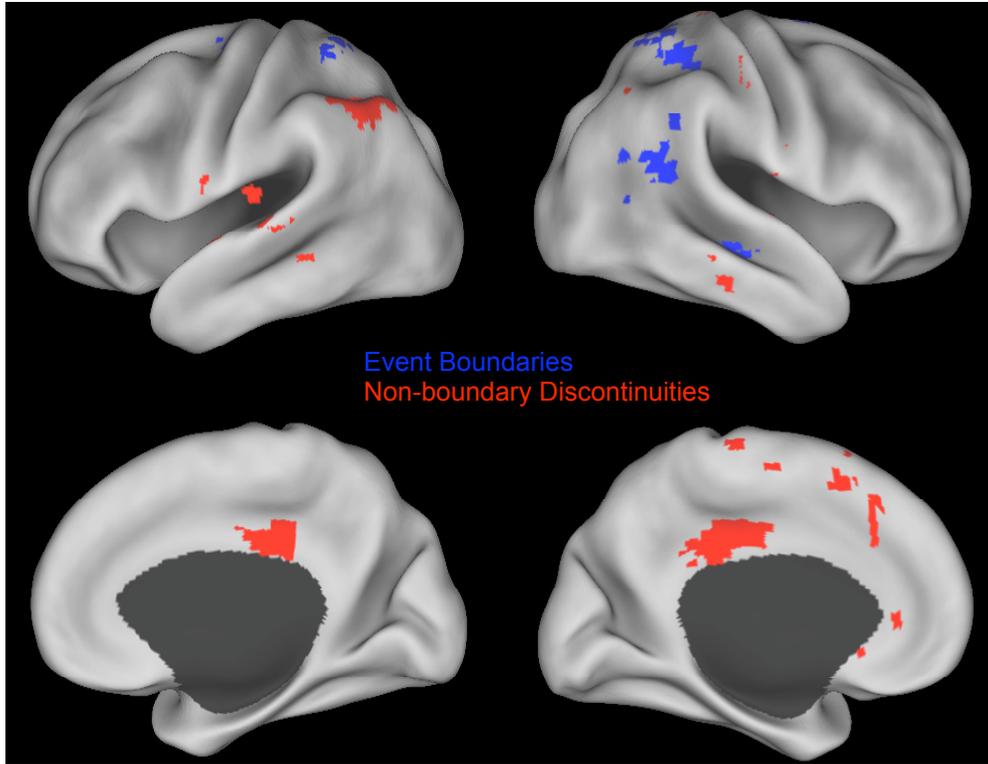


Figure 4. Brain regions that increased selectively at cuts that were identified as event boundaries (blue) or points in time that were discontinuous in space or time but were not identified as event boundaries (red). Data from (Magliano et al., 2007).