# Bridging views in cinema: a review of the art and science of view integration

Daniel T. Levin[1,*] and Lewis J. Baker[2]

Recently, there has been a surge of interest in the relationship between film and cognitive science. This is reflected in a new science of cinema that can help us both to understand this art form, and to produce new insights about cognition and perception. In this review, we begin by describing how the initial development of cinema involved close observation of audience response. This allowed filmmakers to develop an informal theory of visual cognition that helped them to isolate and creatively recombine fundamental elements of visual experience. We review research exploring naturalistic forms of visual perception and cognition that have opened the door to a productive convergence between the dynamic visual art of cinema and science of visual cognition that can enrich both. In particular, we discuss how parallel understandings of view integration in cinema and in cognitive science have been converging to support a new understanding of meaningful visual experience. © 2017 Wiley Periodicals, Inc.

## INTRODUCTION

In the early 20th century, DW Griffith fundamentally transformed cinema by closing the loop between audience responses and the films he created. Griffith was well known for standing in the back of theaters and carefully observing his audience as they collectively inhaled and leaned forward during moments of tension, and subtly sat back when the tension was released. Griffith described this behavior as 'an undercurrent of movement and sound,' and he would often rush into the projection booth to obsessively recut specific scenes in time for the next showing.[1] By iteratively tailoring his films to audience responses, Griffith built upon insights of other early filmmakers to develop many of the principles that transformed cinema from a mere recording and replaying of events into a true art form with its own unique means of expression. Not only do these artist's insights continue to form the basis of modern cinema, but more recently, they have informed cognitive scientists who have mined professional practice in cinema for insights about scene and event perception.[2–4]

We can start with a fundamental problem Griffith faced: He wanted to interrupt ongoing scenes with close shots of actor's faces, and nobody at the time was really sure how to do this without distracting or confusing the audience. Previous filmmakers such as Edwin Porter and George Albert Smith had begun to depart from simply using the camera to record and reproduce events as a whole, but many of these early attempts at within-scene edits were distracting, or involved situation-specific view-linking cues. For example, Smith's film, *Grandma's Reading Glass* (1900) jumps from views of an elderly woman and her grandson to views of what each can see as they look though a large magnifying glass. These between-view edits introduce shots framed by a circular annulus to telegraph the fact that all of the looking views are glances through the magnifying glass. Griffith needed something more subtle—he

*Correspondence to: daniel.t.levin@vanderbilt.edu

[1]Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

[2]Departments of Mathematics and Comptuer Science, and Psychology, Rutgers University, Newark, NJ, USA

wanted to tell stories using close-ups to enhance visual narratives, not to draw attention away from the stories. Audiences had to *see* that each shot continued the story without having to employ deep narrative-disrupting thinking or, worse, having to learn new movie-specific view-integrating skills.

Thus, a key challenge for early cinema was to understand the basic visual-cognitive question: how do people combine disparate views into coherent sequences? This question actually has two important components that will organize this review. First, how do viewers recognize that a sequence of shots depicts a single event? This is a fundamentally cognitive question because it asks whether audiences will be able to integrate the meaning of information in one shot with the meaning of the information in other shots in a sequence. We will refer to these processes as *Conceptual Integration*. The second component of view-combination, *Perceptual Continuity*, is logically distinct from the first. Even if audiences conclude that different shots show the same events, will the sudden cuts between shots induce distractions that lessen audience involvement in the story? At some level, this is a question of aesthetics, but, as we will review below, recent research suggests interesting links between perceptual continuity and conceptual integration that can not only increase our understanding of cinema, but also of event perception more generally.

Questions such as these clearly link the art of cinema with psychology, and early in the 20th century the eminent Harvard professor Hugo Münsterberg argued that cinema was a natural domain for psychological study.[5] Münsterberg suggested that the true power of cinema was its ability to creatively organize a sequence of images and events that escaped the sequential and spatial constraints of live theater and could therefore represent any event, sequence, or juxtaposition that could be imagined.[6] Thus, the cinema is defined by human imagination, or 'mental play' as Münsterberg put it, and the visual narratives filmmakers created could reveal internal mental processes not only by literally depicting characters' thoughts, but also by creating event sequences that revealed the structure of thought itself. For example, Münsterberg (and other early theorists such as V.I. Pudovkin and Lev Kuleshov) proposed that rapid cuts and shifts of camera views could mimic the excited state of a character and thereby induce excitement in the audience.

After this encouraging start at a truly interdisciplinary convergence of art and science, the link between cognitive science and cinema became

relatively dormant for quite some time (with a couple of notable exceptions in the research programs of Julian Hochberg and Robert Kraft). In part this divergence was caused by lack of interest in audience psychology within academic disciplines devoted to the study of cinema. In particular, the discipline of Film Studies was for some time dominated by the argument that socio-cultural forces structure cinema, and therefore that the act of viewing cinema relies more upon enculturation than deployment of natural cognitive and perceptual skills (for review and critique see Ref 7). In addition, progress may have been slowed by the emphasis on simple, controllable stimuli in research on cognition and perception. Recently, however, this chill has lifted not only because of interdisciplinary rapprochement, but also as the result of theoretical developments that support the study of naturalistic events. These developments have produced an exciting flowering of research on the cognitive basis of cinema that not only helps us understand the movies but also can lead to new insights about cognition and perception.

Before describing conceptual integration and perceptual continuity, we want to emphasize that it is easy to become overly focused on between-view integration, and to forget the many other ways in which cinema is a distinctive visual art form. Indeed, important examples of cinema have emphasized long dramatic sequences that contain no edits. For example, the classic Alfred Hitchcock film *Rope* (1948) contains very few edits, and more recently, the films of Alfonso Cuaron such as *Children of Men* (2006) and *Gravity* (2013) feature long takes that depict complex events viewed from a constantly moving camera. Thus, it is possible to deploy a whole collection of cinematic techniques without ever having to disassemble and reassemble different views of a scene. Accordingly we must make clear the limits of this review and refer the reader to earlier research that describes many of these within-frame forms of cinematic practice.[8,9] However, one advantage of focusing on the topic of view integration across edits is that many of the cinematic principles that impact viewers' cognition within shots become most apparent when the filmmaker needs to cut from one shot to the next. At this point the filmmaker must make a crucial decision about what, exactly, is important about the two joined shots, and rely on that judgment to carry the viewer undisturbed (usually) from one view to the next. In a sense, edits are the seams that expose the material of cinema, and so they at least provide a reasonable starting point to understand this dynamic visual art form.

## CONCEPTUAL INTEGRATION

It would be traditional for a review such as this to begin with basic perceptual processes and build on those by discussing cognitive processes. However, we have inverted this sequence to emphasize the importance that filmmakers have placed on conceptual integration in a way that psychological theory may have underemphasized. As described above, filmmakers want viewers both to conclude that different views show the same event, and to experience perceptual continuity. The interesting thing about cinematic editing is that there are circumstances where these two goals conflict, and many filmmakers argue that the primary goal of editing is to meet narrative demands, even when this might induce perceptual inconsistencies (see e.g., Ref 10). On this approach, meaning is the primary determinant of successful view integration. Early support for this view comes a series of experiments by the Soviet Filmmakers Lev Kuleshov[11] and V.I. Pudovkin[11,12] that demonstrated the power of the mind to combine shots across edits. In one experiment, Kuleshov and Pudovkin edited together a series of shots depicting an actor and actress interacting across cuts. The actor looked offscreen to the left and waved, and in the next shot the actress looked back and returned the wave. Then we see the two meet, look offscreen at a building and ascend the steps of the building. Although each of the shots in the sequence were taken in very different locations, this experiment in 'creative geography' produces the impression of a single location by relying upon the audience's knowledge about typical 'greeting' event. Kuleshov in particular, along with his contemporary Sergei Eisenstein,[13] was well-known for arguing that *montage*, the meaningful combination of different views, was the most important property of cinema. Recent research reinforces the importance of montage by demonstrating that jumbling the order of scenes increases detection of events irrelevant to the narrative, presumably due to lessened engagement.[14]

Conceptual view integration relies on a number of filmmaking techniques, each of which may be informed by a closely related domain of psychological research. Generally, editors argue that each new view must be motivated because it answers a question that the preceding shot has induced the audience to ask. One of the most basic questions is induced when an actor looks off-screen, which motivates audience to ask what the actor has looked at. In editing terminology, the sequence showing an actor followed by the thing he/she has looked at is referred to as an *eyeline match*. The well-known director Alfred Hitchcock uses a series of eyeline matches in an iconic scene in *North by Northwest* (1959). As shown in the first frame of Figure 1, Cary Grant has arrived for a meeting in a vast landscape, and he saliently looks offscreen, inducing the audience to ask what he has looked at. The target of his gaze, an ominously empty road, is shown in the next shot. This question-answer pattern is often followed by a shot revealing the character's reaction to the newly seen information, as demonstrated by Cary Grant's puzzlement in the third frame of Figure 1. Thus, the director trusts first that audiences will follow Grant's intention to look offscreen, second that they will understand how the eyeline match shows what Grant has looked at, and finally that they will more deeply consider Grant's mental state in reaction to what he has seen.

The ability to track gaze and to consider the beliefs of the looker mirrors theories explaining the development of mental state reasoning. During their first year of life, infants learn to look where an interactive partner is looking, and by the end of their first year, infants can follow gaze to objects that were initially out of their view or hidden by a barrier.[15–17] This likely reflects a very simple form of attentional spotlighting that leads infants to look at important things. Developmental researchers argue that this basic spotlighting is soon elaborated into an understanding that looking is associated with intentional mental states such as beliefs, desires, and goals.[18]



**FIGURE 1** | This eyeline shot in *North by Northwest* demonstrates two cognitive functions that develop with age and experience. Cary Grant looks off screen in shot (a), and we understand he is looking at the farmland in shot (b). The viewer is prompted to elaborate on Grant's mental state in shot (c).

More recent research demonstrates that many of these skills are part of a default gaze-tracking system in adults,[19,20] making them excellent candidates for simple default cognitive skills that can be used to integrate eyeline matches and reaction shots.

Over their first few years, children build upon these basic understandings of gaze and mental states, leading to the development of a *theory of mind*, the ability to understand how another person's perspectives and intentions guide their behavior.[21] For example, one signature of a fully developed theory of mind is the understanding that people often operate under the burden of *false beliefs* that guide them to do one thing because they are unaware of facts that would lead them to do something else. The ability to understand false beliefs develops considerably later than more basic forms of gaze following and belief attribution, and several recent theories of mental state reasoning propose a distinction that persists into adulthood between basic processes that are invoked by default with relatively little cognitive cost, and the deeper, nondefault capacity-requiring belief attributions necessary to track false beliefs.[22,23] Research reinforces the hypothesis that these deeper forms of mental state attribution do not occur by default and require capacity by demonstrating that adult participants often fail to engage them, especially when participants are under cognitive load or time constraints.[24] These nondefault forms of mental state reasoning are nonetheless a useful a tool for engaging audiences with a narrative,[25] and with cinema[26] . Some of the most interesting evidence supporting the utility of theory of mind in cinema comes from evidence of unusual brain activation patterns during film viewing in individuals with autism-spectrum disorders (ASD) who are thought to exhibit a theory of mind deficit.[27] This research builds upon previous findings demonstrating that when different viewers see a film, their brain activations are highly correlated, revealing the common perceptual and conceptual response to the film.[28] This intersubject correlation extends across many visual areas, but includes higher-order attention and association areas as well. However, individuals with ASD show fewer intercorrelations,[29] suggesting that their lack of theory of mind skills leaves them relatively adrift when interpreting cinema.

Mental state reasoning likely underlies the affective experience of cinema as well. For example, viewers' experience of suspense has been directly tied to their interpretation of characters' mental states.[30] Researchers have also documented how knowledge generates suspense[31] and how cinematic suspense in turn organizes visual scene processing by increasing brain activation in early visual regions responsible for central processing, and in a parietal attention network, while decreasing activations in peripheral processing regions and in default mode regions responsible for non-task-directed cognitions.[32] Other work has explored how 'fictional world' emotions (e.g., sympathetic sadness in response to a character's depression) and 'artifact' emotions (e.g., fascination and anxiety) increase immersion in a filmic narrative.[33]

Of course, not all edits are organized by interpretations of looking, so in addition to gaze following and mental state reasoning, conceptual view integration relies on broader skills that support event perception. It is particularly interesting to note that psychological research exploring this domain has been strongly influenced by cinema. Event perception research has primarily focused on how perceivers organize continuous action into discrete meaningful units. In a foundational study of event perception, participants watched a film and pressed a key every time they judged a new event to be beginning. Not only did viewers generally agree where events started and ended, but their recognition memory for film stills was best at the boundaries between two events.[28] Related research suggests that audiences store relevant information about the current event in working memory that is partially or fully refreshed at the instantiation of a new event.[34,35] Findings such as these were later incorporated by Jeff Zacks into Event Segmentation Theory,[36] a model of event perception whereby encoding and memory of visual actions is organized by abstract (and relatively sparse[35]) event models held in working memory. One key feature of event segmentation theory is that viewers can choose to focus on events at a relatively coarse grain, reflecting abstract goals, or at a relatively fine grain, reflecting simpler actions. These different levels of grain are represented hierarchically such that fine grain events are represented as components of coarse grain events, and all of these levels of abstraction are maintained in an active state while events are occurring.

Another key feature of event segmentation theory is that event segments are caused by a forward prediction process that generates predictions based on the current event, and when the predictions are incorrect, a prediction-error signal induces increased perceptual encoding and event segmentation.[37,38] This emphasis on forward prediction can be seen as an interesting contrast with models of narrative comprehension during reading that emphasize bridging inferences. Bridging inferences derive the meaning of current sentences from the immediately preceding

text[39] and theories of narrative comprehension often emphasize how forward prediction is limited to situations that demand it (for review, see Ref 40). However, one interesting possibility is that readers make fewer predictive inferences than film viewers because cinema relies upon editing and framing devices that provide visual support for predictions. In one test of this hypothesis, researchers observed that film viewers made predictive inferences much more frequently in the presence of cinematic cues such as framing, editing, and character actions, although in all situations, substantial numbers of viewers did not show evidence of predictions.[40] Generally, this is consistent with the theory of mind research demonstrating that deeper consideration of the beliefs that organize actions is a nondefault limited capacity process. In addition, recent work testing the perception of short films reveals the impact of simpler and more readily generated inferences about immediately previous events. In these experiments,[41] participants falsely recognized a still frame depicting the moment someone has kicked a ball even though the film they saw showed only the prekick windup and then the ball bouncing down the field. False recognition was less likely when the subsequent ball-bouncing shot was omitted, implicating its role in generating an inference about the previous kicking action.

Collectively, these results suggest that film viewers rely both on predictions and inferences about the past to identify and understand visual events. To understand these processes, it is important both to specify broad categories of cues that might change to indicate a new event, and to contextualize predictions and inferences in the broader narrative structures that contain them. In pursuit of the former goal, extensive research supporting both event segmentation theory and theories of narrative comprehension specify dimensions such causality, time, space, characters, and goals[42] as the basis for these forms of integration, and links between these dimension changes and event segmentation has been confirmed for feature films[43] and television sitcoms.[38] Neuroimaging research at least tentatively supports these observations as well, demonstrating that brain responses evoked by event boundaries could be predicted in large part by changes in these dimensions[43] and that edits in films that induce spatiotemporal changes are associated with increased activation in areas that encode spatial locations.[42] The second goal of contextualizing event perception in larger narrative structures has been met by a body of theory describing how typical cinematic narratives can often be parsed into a set-up, complication, development and climax that contain characteristic events.[44]

Recent statistical analyses of shot length and between-shot transitions demonstrates average shot length increases at boundaries between these large segments.[45] Finally, recent brain imaging data suggest that distinct forms of memory integration occur across timescales less than, and greater than, several minutes. In particular, integration across minutes appear to be supported by joint action of default mode regions and the hippocampus, while longer timescales are more exclusively the province of hippocampal activity.[46]

If conceptual integration in editing relies on fundamental cognitive processes, then it would follow that even individuals with no experience with cinema could draw upon these processes to integrate views. A small number of fascinating cross-cultural studies partially support this claim, although there do seem to be interesting limits to first-time viewers' understanding of cinema. Early work demonstrated that individuals from a remote African tribe who had never seen cinema showed no deficit in understanding a narrative about the theft of a cow when it was presented using a series of shots joined by cuts relative to when it was simply presented in one long take.[47] More recently, Sermin Ildirar has been testing a population of individuals in the mountains of Isparta, Turkey, who have not seen any edited films. Initial reports[48] suggested that these individuals showed a mixed pattern of success. For example, the first-time viewers were shown a simple 'shot-reverse-shot' sequence in which successive views each showed one actor sitting still, looking at the other as if they were about to speak (Figure 2). Nominally, this shows two people sitting in one location looking at each other, and this is how the control group of more experienced rural Turkish viewers interpreted the scene. The first-time viewers, however, did not integrate the two views, interpreting them as depicting as two people who were not interacting. However, these same viewers had little difficulty interpreting parallel action in which views of one familiar activity were intercut with views of another simultaneously occurring familiar activity, and a subsequent experiment demonstrated that first-time viewers understood a sequence in which actors exchanged an object across an edit.[49] Recent work with this population has demonstrated that first-time viewers correctly interpreted the shot-reverse-shot sequence when the actors actually exchanged a simple familiar greeting (S. Ildirar, unpublished data).

One interesting interpretation of this cross-cultural work is that the cinema-naïve viewers can rely on familiar events to integrate views, while the more experienced viewers did not need this, and

**FIGURE 2** | Example of reverse angles from Ref 48. Reverse angles can be seen as a special version of a repeated set of eyeline matches in which an actor looks offscreen at another actor who appears to look back at the first actor. These reverse angles are slightly atypical as filmmakers often use views that are not directly perpendicular to each actor and look more into the actor's faces. These are stills from a film in which the actors did not interact. Naïve viewers did not integrate these views. However, when the actors spoke or engaged in an identifiable action, naïve views could correctly integrate the views.

likely interpreted the scenes by drawing upon more general knowledge about pragmatic goals of cinema as a medium. These pragmatic intuitions are an important component of media understanding and they are the focus of developmental research exploring children's understanding of the referential nature of cinema. For example, research documenting the 'video deficit' shows that young children understand how video depicts meaningful events but have difficult realizing that video is not always a cinematic representation of nonpresent fictional worlds and can sometimes represent the here and now.[50] This limit reveals itself in children's striking failure to use a video of an object being hidden in an adjacent room to locate that object. More generally, the pragmatics of cinema may reflect a variant on people's understanding of communicative intentions in language (e.g., the understanding that an utterance is intended as a means of transferring ideas[51]). It is interesting to consider the possibility that although the basics of conceptual integration may rely on skills that are easily applied to cinema, and therefore do not require much medium-specific learning, learning is necessary to achieve a deeper appreciation of how cinematic communicative intention can range from depicting fictional events, to communicating abstractions, to recording real-world events.

## CONCEPTUAL INTEGRATION IN ALFRED HITCHCOCK'S THE BIRDS

In this section, we have reviewed research, theory, and filmmakers' more informal observations that not only highlight the importance of meaning in integrating views, but also give us a good sense of how this integration occurs. If one were to summarize the

research in this section, it might be to state that view integration is structured by knowledge about social agents, and events more generally. These processes allow viewers to make predictions about the future and to integrate knowledge about the recent past. To get a sense for how this might work, and to discover ways in which cinema might expand our understanding of view integration, consider a suspenseful scene from Hitchcock's well-known film *The Birds* (Figure 3) in which the character Lydia explores the disturbingly silent house of her neighbor. Prior to this scene, Lydia has just experienced an evening of terror in which hundreds of crazed birds have invaded her own house, attacked her family, and broken things. In the aftermath of the previous attack, Lydia had been shown moving among her family picking up broken china (Figure 3(a)). Lydia's dissipated handling of the broken china is reinforced by the concerned gaze of her son's (potential) girlfriend, making clear Lydia's trauma. Despite her upset, Lydia decides to check on her neighbor, and in the depicted scene, suspense builds as Lydia enters his creepily quiet home and notices the broken teacups in his dining room. Lydia, is horrified to see this, and pauses (Figure 3(b)). At this moment, the viewer must rely upon most of the processes described above when integrating this view with the next one showing Lydia in the hallway as she continues to search the house. The viewer's understanding that the shot of Lydia in the hall (Figure 3(c)) is the next step in Lydia's search starts with basic gaze-following processes that spotlight the broken cups and mental state attributions that support the inference that Lydia is thinking about the cups. This is elaborated by deeper theory of mind skills that lead viewers to detail Lydia's specific beliefs about the cups, which in turn requires narrative inferences about the recent past. These inferences are used

**FIGURE 3** | Conceptual integration in Hitchcock's The Birds. In order to integrate the middle view with the narrative, viewers must consider how the character's fear was conditioned by her recent past (as represented in the left-hand image which shows her picking up broken cups from a previous horrifying bird attack). This information must be used to project into the future as viewers consider whether she will continue to explore the apparently empty house, or will run away (she eventually does both).

to generate predictions about Lydia's behavior that diverge in interesting ways from the predictions hypothesized by event segmentation theory. In particular, it appears characteristic of a suspenseful event that audiences internally represent two contrasting predictions—that Lydia will run out of the house *and* that she will continue into the hall. Thus, it is possible that this instance of cinematic view integration suggests an extension to theories of event perception whereby representations of several possible next events are not only compared with the eventual input, but are also compared with each other, and with the recent past.

It is very interesting to note all of the pains Hitchcock took in the preceding scene to emphasize Lydia's interaction with the cups, and this may embody the director's understanding that many theory of mind inferences do not occur by default, but can be supported by meaningful visual events. At least one review of the role of theory of mind in literature speculates that visual support allows a deeper level of theory of mind reasoning than would otherwise be possible.[25] Empirical evidence for this hypothesis comes from a range of sources including classic research documenting that meaningful between-object visual relationships improve memory[52] and that animated visual diagrams support insight problem solving.[53] However, visualization does not always facilitate problem solving, especially if it is ineffectively targeted,[54] which suggests the utility of drawing upon Hitchcock's apparent wisdom in shaping the audience's reasoning about Lydia's beliefs. In particular, the scene appears to embody the hypothesis that a full complement of suspense, forward/backward inference, and dramatic gaze

shifts are necessary for visual events to support theory of mind reasoning, and that simpler depictions of spatial relationships between characters are not sufficient.

Another avenue for research suggested by this example might be to explore how the ebb and flow of forward prediction and backward inference varies in different event domains. In the Birds example, predictive processing reflects a very elaborate set of links between the current event and both the past and the future, but other settings might be characterized by a different mix of these processes, or by far fewer of these cognitive linking processes overall. This possibility is reinforced by a recent theory of executive functioning that hypothesizes a contrast between a proactive mechanism that continually uses goals to shape processing of incoming information, and a reactive mechanism that only invokes goal-related processing when events demand it.[55] On this view, suspense scenes such as the Birds might be highly proactive, while other kinds of scenes such as a chase or an abstract montage might be more reactive. In addition, consideration of cinematic events begs the question of whether these mechanisms can trade-off within scenes, or whether shifts between these mechanisms might be signals for new events in themselves.

Finally, this example foreshadows the key issues of perceptual continuity that will be covered in the next section. The scene includes a large number of visual properties, spatial relationships and events such as movements, transients between cuts, and small actions. The key question is, how does the visual system make sense of the scene by intelligently tracking some of these, while remaining unaffected

by others. Clearly, the meaningful structure of the scene is not the whole answer to this visual selection and integration problem, but it is part of the answer, and so the next section explores perceptual continuity both by considering the role of basic perceptual processes in cinema, and by considering how these processes support (and are supported by) meaning.

## PERCEPTUAL CONTINUITY

As reviewed above, filmmakers are concerned that audiences understand how a collection of views can be integrated into a meaningful event, but also that this integration proceed without distraction caused by sudden shot-to-shot transitions. In a way, the issue of perceptual continuity is simple—edits introduce sudden appearances, disappearances, apparent movements, and property changes, and filmmakers usually want these to attract as little attention as possible. Some of these sudden changes produce perceptual transients that have the potential to attract attention by activating very basic visual event detectors. For example, a cut from a long shot showing a person standing in the desert with a mountain in the background to a close-up of the person's face will cause the mountain to suddenly disappear while the large-scale image of the face appears. A different issue arises when one cuts between more similar views—if one suddenly cuts from one view of a face to a different view of the same face depicted from a slightly different angle in a slightly different location, the faces will be similar enough to induce long range apparent motion, and the face will appear to rotate and move.

In practice, the issue of long-range apparent motion is much more problematic for filmmakers than the more basic issue of attention-grabbing appearances and disappearances.[56] Filmmakers rely heavily on the supposition that cuts can be unobtrusive if they avoid apparent motion by juxtaposing substantially differing views,[8] and recent experimental evidence demonstrates that many of these edits go undetected even when viewers are purposely looking for them.[57] This phenomenon, referred to as *edit blindness*, converges with other research demonstrating that viewers often fail to detect disruptions to videos in which the screen is blanked for up to 600 ms.[58] Both of these findings are particularly interesting in light of research on attentional capture, which initially hypothesized that a limited set of perceptual events such as appearances and disappearances attract attention in a bottom-up stimulus driven fashion.[59] This understanding of attentional capture would appear to preclude filmmakers' ability

to make unobtrusive cuts. More recently, though, the initial strictly stimulus-driven capture hypothesis has been revised by data demonstrating that capture is strongly modulated the attentional focus of the observer.[60]

Although these more context-sensitive forms of attentional capture would appear to accommodate unobtrusive edits, most of the laboratory tasks used to demonstrate this flexibility are quite unlike cinema. For example, many tasks purposely draw participants' attention to a particular capture cue (such as color or onsets) in the moments just before that cue becomes a candidate to capture attention.[61] In contrast, we suspect that the more general focus on the meaning of events is the key factor that lessens capture in cinema. This would be generally consistent with recent work documenting the impact of conceptual relatedness on capture,[62] but instead of focusing on one aspect of a stimulus to modulate capture in another, cinema focuses attention on meaning at the expense of *all* perceptual transients. This form of attentional modulation would be compatible with models of meaning in event perception such as Event Segmentation Theory, which emphasizes how event models held in working memory guide attention by selectively gating perceptual input during event boundaries.[36]

In fact, the failure of transients to consistently attract attention allows them to hide other potentially problematic between-view inconsistencies in object properties, configurations, and locations. When filmmakers first started editing, they were naturally concerned that even small changes in body position, costume, and background would grab viewer's attention. However, even by 1920, when continuity editing was in its infancy, the Soviet Filmmaker Lev Kuleshov noted that small changes to actors and their costumes often escaped notice.[11] Observations such as these are the basis of current editing practice that attempts to avoid large perceptual changes, while allowing smaller changes, and sometimes even leveraging the flexibility afforded by these loose expectations.

It is interesting to note that by the 1980's, the dominant psychological hypotheses explaining view integration depended on the assumption that high-capacity visual buffers were necessary to integrate views by comparing and registering visual properties between views. However, even initial arguments against the integrative buffer hypothesis pointed out how it relied too heavily upon artificial tasks,[63] and early work documenting change blindness[64] drew upon the insights of filmmakers to demonstrate the extent to which viewers fail to detect visual changes

that occur during edits. In one experiment, participants viewed a short film that used common cinematic editing practice to depict two actresses conversing. Every edit in the film included a sudden change in object properties (e.g., an actress's scarf disappeared, a hand position changed, and plates on the table in front of the actors changed from red to white). Even in the presence of so many perceptual inconsistencies, participants easily integrated the views into a coherent-appearing event, and they never noticed the changes unless they were specifically looking for them (in which case they still only noticed a small proportion). More telling, even changes in attended objects often escaped notice—when the sole actor in a scene changed across a cut from one person into another most participants still failed to see the change. These failures can be seen as evidence that view integration does not depend on moment-by-moment tracking of perceptual details.

Failures to detect changes, referred to as *change blindness*, have been observed in many contexts, ranging from still images of natural scenes to arrays of natural and artificial objects (for review, see Refs 65,66). Related phenomena demonstrate similar failures in awareness of nominally salient stimuli and events. *Inattentional blindness* makes clear the power of viewers to maintain awareness of one object in a dynamic scene while remaining unaware of other objects,[67] even if they are bizarre or noxious (for example a Gorilla walking through the scene;[68] or a woman walking into the scene and scratching her fingernails on a chalkboard[69]). *The attentional blink*[70] and *repetition blindness*[71] reveal how awareness of an object at one moment in time can limit awareness of subsequent objects.

Several ideas from this blindness-documenting literature are important for present purposes. First, these studies reveal an interesting balance among the potentially attention-attracting changes and transients that occur in edits. For example, sudden onsets similar to those that occur during edits can induce change blindness,[72] and recent work has demonstrated that such transients can guide attention even in the absence of awareness.[73] So, it appears possible that the transients induced by edits leverage a form of sub-threshold noise that can lessen the detectability of other sudden changes. Second, these failures of awareness are not limited to mediated representations of objects and events—they also occur during real-world interactions. Change blindness has been documented for sudden swaps of real-world conversation partners,[74–76] and inattention blindness has been documented in real-world simulations of crimes.[77] This commonality is important because it is

evidence against the hypothesis that failures to detect changes and other transients are unique to cinema. Therefore, research documenting basic constraints on awareness is relevant to cinema and, conversely, perceptual phenomena observed in cinema may be relevant for understanding visual cognition.[2] Third, many of these findings are responsive to forms of knowledge and task contexts[49,78] that affect visual attention. For example, change blindness is lessened for objects that are central in a scene,[79] or are distinctive in a given scene context,[80] and for objects that are associated with an expert's knowledge about a scene.[81,82]

When making the link between between perceptual view integration and meaning, it is helpful to consider the other side of the between-view integration question: If transients and property changes do not consistently produce discontinuity, is any sort of between-view change reliably disruptive? Long-range apparent motion clearly interferes with continuity, but even if this is avoided, one cannot simply cut any two views together and expect a smooth transition. In addition, even if one accepts that property changes and transients are often missed, one still has to explain the substantial proportion of situations in which these inconsistencies are disruptive. To encompass both how different kinds of changes and how specific situations modulate continuity, Smith, Levin, and Cutting[3] proposed a *focused continuity* hypothesis, which states that visual continuity relies upon only a subset of possible continuity-supporting cues, and that viewers are sensitive to variations in the reliability of different cues for any given scene or event. Although property changes and transients often fail to interfere with continuity, there may be occasional cases where these inconsistencies do become important (as suggested in Figure 4). Other cues may be more consistently important in establishing continuity, but even in these cases, viewers may be able to vary their dependence on them. The most well-studied of these cues are spatial—filmmakers have observed that violations of basic spatial relationships (usually between actors) can produce an attention-grabbing form of disorientation, as if actors have suddenly moved, or are looking in the wrong direction.[41] To avoid these disruptions, filmmakers rely on the 180 degree rule which specifies that all camera positions should be constrained to one side of the two currently most salient centers of attention. As illustrated in Figure 4, shots consistent with the 180 rule depict actors who consistently look off-screen at the other actor, whereas camera positions from the other side of the line result in shots in which an actor appears to look at nothing. Research has
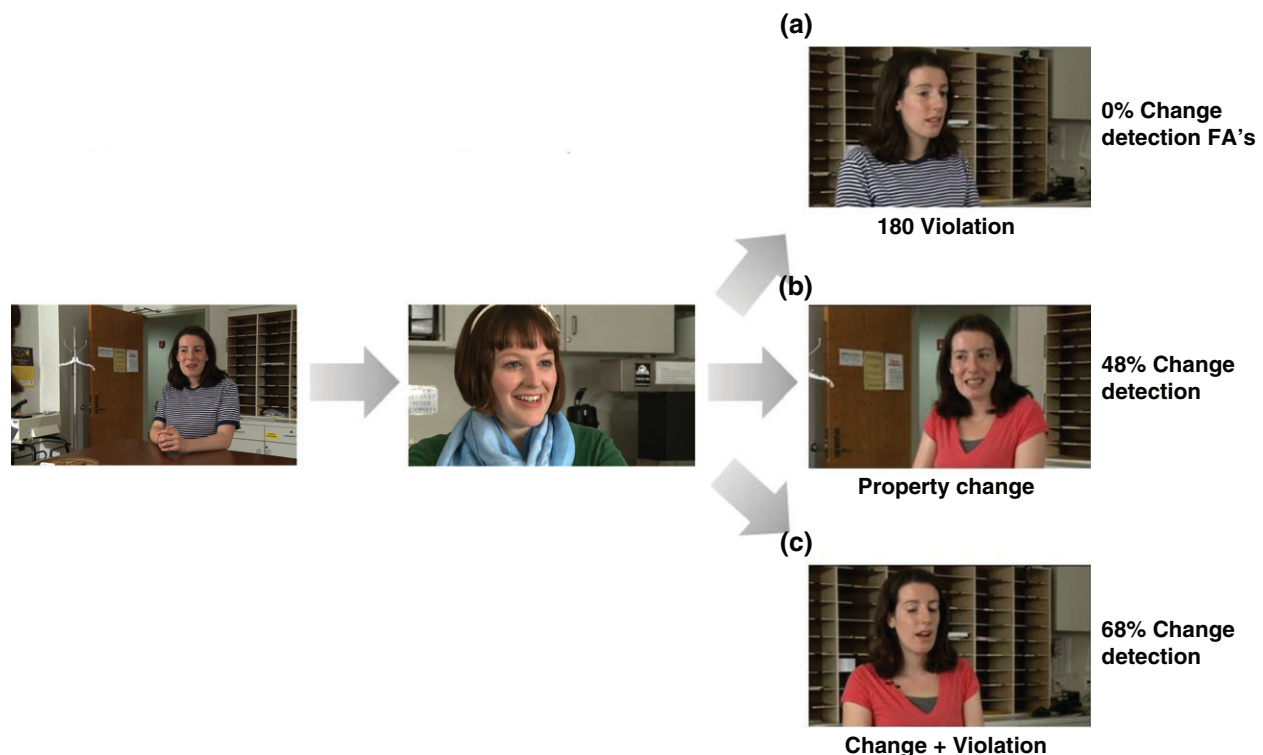
**FIGURE 4** | The 180 rule, and its interaction with change detection. The 180 rule is often invoked to maintain perceptual continuity when filming reverse-angle conversations. According to this rule, the camera must stay to one side of the conversing actors, which results in a series of shots in which each actor appears to look off-screen at the other. When this rule is broken (as in still (b)), the actors appear to be suddenly displaced. In a recent experiment, this displacement induced comparisons between properties in working memory, and the current scene, significantly increasing change detection for the substitution of the actor shirt.

demonstrated that sequences violating the 180 rule are less accurately remembered,[8,83] that the spatial layouts they depict are less accurately reconstructed than equivalent nonviolating sequences,[67] and that violations impose a reaction time cost on spatial encoding in animated scenes.[84] Findings such as these are consistent with the hypothesis that spatial encoding is one of the most important bases of cinematic continuity.[9,85] It is, however, interesting to note that no research has carefully assessed incidental detection of 180 violations (although participants in one recent study did detect about half of violations without forewarning, the study was not designed to carefully assess this[86]), and so for now this remains a possible target for some potentially interesting new research.

Despite the lack of empirical confirmation that 180 violations produce perceptual discontinuities that reach awareness, the fact that these disruptions interfere with spatial representations and event memory suggests that this form of spatial encoding occurs by default when viewing events in cinema. If so, cinema's use of this principle may prove informative about spatial representation during real-world event perception. The first interesting observation one can

make about the 180 rule is that it appears to structure basic spatial representations around meaningful gaze cues. Thus, this may be an important case where perceptual continuity and conceptual view integration meet. Also interesting is the fact that filmmakers do not always adhere to the rule. Exceptions are well known, and a recent small-scale survey of modern films and found that approximately 10% of sampled scenes contained at least one edit that did not adhere to the 180 rule.[87] The few violations were very informative. One thing that many had in common was that they preceded the violation with shots in which the spatial axis of the larger scene was very salient. For example, in one violation from the movie *Lord of the Rings: Return of the King* (2003), the shots bracketing the violation clearly specified the spatial geometry of the inside and the outside of the cave the characters were standing in. This implies that filmmakers intuit that viewers can flexibly select from among a set of spatial cues to integrate views, depending on which is more useful in a given scene. This converges with findings from the spatial cognition literature not only because this literature has emphasized the importance of similar cues such as

spatial geometry,[88] but also because recent theory in spatial cognition hypothesizes that perceivers adaptively derive representations from different cues, depending on each cue's reliability in the current context.[89] Thus, basic research in spatial cognition can be seen as providing support for the focused continuity hypothesis, and, in a more basic way, observations of cinema add to evidence suggesting that at least some forms of spatial encoding occur by default.[90–92]

Given the possibility that the 180 rule reflects a form of default spatial encoding, the rule may be a useful lever for understanding the role of short term spatial encoding during event perception. One interesting question is whether these on-line default encodings eventually lead viewers to generate fuller detailed representations of spatial environments. The narrative-encoding literature suggests that detailed spatial representations are created only when readers feel the need to encode them.[93] For example, participants often derive little spatial information from a text unless they are explicitly instructed to focus on it.[94] Consistent with this idea, Hochberg and Brooks[95] argued that only local forms of view integration depend on spatial representation. However, other work does argue that at least some forms of long term visual representations are accumulated by default,[96] and Magliano, Miller, and Zwann[97] argue that visual narratives such as cinema support the creation of spatial representations. The prevalence of longer-term spatial representations in cinema was tested behaviorally by asking fans and nonfans of well-known television series to make spatial judgments about locations on the sets of the shows.[98] In some cases, experienced viewers were very good at judging the relative displacements of different locations on these sets, but in others, viewers were no better than nonviewers in making these spatial judgments. The factor that differentiated these two findings was that sets with a 'fourth wall' were effectively represented, while sets that surround the viewer were not represented. For example, the Drew Carey Show is filmed from a fourth wall where audience sits, so the kitchen is always screen-right from Drew's living room. In contrast, some shows are shot in sets that surround the camera. This means that a given location on the set could be screen-left or screen-right relative to another location. So, on the set of the television drama ER, if one is looking from the inside of the ER toward the main entrance, most of the rest of the hospital is to the right, but sometimes, the camera looks the other direction, from the entrance to the inside of the hospital, and in this case the rest of the hospital is screen-left. Thus, it appears as

though viewers do build a mental map of depicted sets, but only when relatively simple screen-centered directional encodings are sufficient. On the assumption that screen-centered and viewer-centered reference frames are similar, this finding may find support in research demonstrating that the parahippocampal place area (PPA), a brain region that is associated with encoding spatial environments, relies on a similar viewer-centered reference frame.[99]

If the default spatial encodings represented by the 180 rule result in mental maps under only limited circumstances, one might reasonably ask what purpose this form of encoding serves. Recent work in our laboratory suggests that the answer may reflect how spatial changes induce perceptual comparison processes that support event perception. As noted above, one of the cues that induce event perception is a spatial change. Thus, it is possible that 180 violations are disruptive because they mimic a spatial reconfiguration that would normally signal a new event. Based on this reasoning, Baker and Levin[86] hypothesized that 180 violations would induce visual-cognitive processes such as between-view property comparisons that prepare for identification of new events. In a test of this hypothesis, participants viewed short edited films in which between-view visual property changes coincided with edits that did or did not violate the 180 rule (see Figure 4). Consistent with the idea that these violations would induce property tracking in preparation for encoding a new event, violations were associated both with increased change detection and with increased judgments that new events had begun. In addition, new events marked by meaningful goal changes also induced increased change detection. A key feature of this finding is that change detection always required a comparison of a currently visible feature with a feature held in working memory: the prechange feature was only visible in a previous shot that had not been on-screen for several seconds by the time the postchange feature was present (see Figure 4). So, this work converges with recent research suggesting that visual and linguistic working memory systems play an important role in on-line inference generation when viewing brief visual narrative events.[100] It is also compatible with work suggesting that event boundaries are associated with increased elaborative processing that lessens visual probe detection.[101]

This work suggests that both the meaningful structure of events, and more basic spatial cues associated with events can impact property comparisons across cuts. Related work asks whether other forms of event-based encoding occur as a perceptual default in the service of event understanding. For example, if

an actor taking a drink has moved the cup 50% of the distance toward their mouth in one shot, should an edit to the next shot start with the cup slightly more advanced in position, exactly where it would be if the shot had picked up 1/30 of a second later? Film editors are divided on the issue, with some arguing that such a match is best, while others argue that a few frames of action overlap is necessary to allow the mind the time to 'catch up' with the new view.[10] Only a few studies have assessed this in a limited range of circumstances, but initial results suggest that a small overlap looks best to viewers.[102] However, preliminary research in our lab suggests that this preference occurs only when participants scrutinize multiple versions of edits, and that otherwise there is very little difference in the perception of smoothness across amounts of overlap and ellipses of up to 12 frames (L. Baker, & D. Levin, unpublished data). Results such as these suggest relatively few default expectations about the exact position of objects across edits.

If inconsistencies in object position do not interfere with perceptual continuity, and basic event knowledge induces perceptual filling-in inferences across shots (as in the research reviewed above in which participants imputed an unseen 'kicking' event), do other forms of knowledge about events act as default integrative cues? In particular, recent work on the comprehension of visual narrative hypothesizes that encoding the sequence of events is necessary for (and prior to) comprehension,[103,104] and as mentioned above, Zacks's Event Segmentation Theory posits a continuously running predictive process that produces an error signal if a given event is not followed by an expected subsequent event.[37] Both of these ideas at least imply that violations of correctly understood event sequences should be readily detected. However, recent research suggests that detecting sequence reversals in edited films depicting event sequences is quite difficult[105] For example, in one film, a close up of an actor's hand grabbing a screwdriver was shown *after* the actor could plainly be seen using the screwdriver. The reversed sequences were 1–2 seconds in duration. Even so, participants had considerable difficulty detecting reversals when they devoted their full attention to the task (see also Ref 106), and when attention was divided, or when participants were not warned to detect reversals, detection essentially fell to chance levels. These results suggest that perceptual sequencing of actions and events across views does not appear to be a capacity-free default process, perhaps because it requires substantial effortful integration between currently visible events and abstract event models. It is

particularly interesting to note that in contrast to change blindness, filmmakers do not appear to have discovered viewers' flexibility with regard to sequence.

Collectively, these results suggest that knowledge about brief actions (one might classify them as subevents) structures perception across edits in as much as they facilitate the identification of larger events, but this identification is not based on detailed tracking of either the spatiotemporal or sequential information characteristic of those actions. This relatively abstract form of on-line tracking may be characteristic in particular of cinema where narrative understanding, not visual coding, is the goal.[55,105,107]

## IMPLICATIONS OF PERCEPTUAL AND CONCEPTUAL CONTINUTY FOR VISUAL ATTENTION

According to the focused continuity hypothesis, view integration across cuts depends on selective subsets of possible view-linking cues, and these range very broadly among basic forms of apparent motion, between-view property matches, various spatial cues, and meaningful event structure. Because cinema viewers are unlike laboratory participants in that they are under no obligation to process visual information for its own sake, the amount of actual visual information they process can fall particularly low ebb when these contextual factors do not demand it. This implicates a key role for visual attention in selecting only the small subset of visual information necessary to extract cinematic meaning. It is important to emphasize that many of the issues described above are often discussed in the context of visual attention, while our review has focused instead on the specific forms of information that are targets of attention, and on the consequences of attending. However, other recent reviews of the science of cinema have focused more directly on visual attention itself. Recently, Smith's Attentional Theory of Cinematic Continuity[108] proposed that the goal of editing is not to create a rich representation of a scene but rather to support shifts of visual attention that efficiently guide processing to informative regions of the screen at useful moments. Generally, this is consistent with much research on attention during visual events in that it emphasizes how attention selects or inhibits one bit of visual information at the expense of another. However, in the case of cinema, where visual information is organized by a deeply meaningful narrative, attention has another equally important role to play. It has to intelligently mediate between

elaboration on visual information and focus on internal representations of events. At some moments, it is necessary to elaborate on specific visual information, but if these moments are chosen inefficiently, processing of sensory information may result in failures to devote sufficient resources to the mental models (such as event models) and memories that are necessary to interpret that information. A similar trade-off can be seen in attentional-blink-induced failures to process important visual information that has followed pointlessly attended unimportant visual information.[109] This view converges with recent theories of attention that emphasize a dichotomy between externally and internally directed attention,[110] along with research emphasizing importance of understanding shifts between these two forms of attention.[111,112]

## PERCEPTUAL CONTINUITY IN ETERNAL SUNSHINE OF THE SPOTLESS MIND

One of the most interesting ways of observing the role of perceptual continuity in cinema is to focus on situations where filmmakers purposely violate it, in part because these violations can create meaning in ways that cognition researchers may not have fully considered. Perceptual continuity violations are common in the film *Eternal Sunshine of the Spotless Mind*, directed by Michele Gondry. The film tells the story of Joel and Clementine, a couple who, in the misery of their breakup, have rashly decided to undergo a neurological memory-erasing procedure. The film begins as Joel discovers that Clementine has erased him from her memory after a fight, and so he decides to do the same. However, once the procedure has begun, Joel has second thoughts, and we journey

between reality and depictions of Joel's memory as he wages a desperate battle within his own mind to subvert the memory-erasing procedure and to retain a trace of his knowledge of Clementine.

Figure 5 shows stills from two scenes, both of which include purposeful continuity violations that create a cinematic depiction of the malleability of Joel's internal memories. The winter beach scene shown in stills 1–4 depicts Joel's disappearing memory of an off-season at the beach with Clementine. We see the couple running and playing in the snow, but we are quickly alerted to the fact that the scene depicts a memory by a series of perceptual continuity violations in which camera angles change minimally during cuts, producing a strong exogenous percept of apparent motion. Stills 1 and 2 are successive frames separated by a cut, as are stills 3 and 4. The strong graphic similarity between stills 1 and 2 allow the audience to experience an ellipsis that simulates Joel internal mental battle to retain his memory, and the sudden disappearance of Clementine from Joel's clutch induced by the cut between stills 3 and 4 creates a visceral perceptual experience of loss.

The scene depicted in stills 5–8 is, in interesting ways, the converse of the first scene. The scene depicts Joel and Clementine walking on a busy street. At first, the couple chats amicably, but they quickly veer into an argument. The beginning of the scene is characterized by traditional continuity editing and nothing is amiss, but as the argument begins, discontinuities appear with increasing salience. For example, one of the initial discontinuities is the disappearance of the printing on the sign above Clementine's head in stills 5–6. The disappearance does not coincide with a cut, but is instead a fade-out over nine frames (about 300 ms). Soon thereafter, a strange faceless being walks by Clementine to her right (in still 6). This time, the oddity



**FIGURE 5** | Intentional perceptual discontinuities in Michele Gondry's *Eternal Sunshine of the Spotless Mind*. Note that stills 5–8 are enlarged by approximately 50% to highlight changing details.

is accompanied by a buzzing sound reminiscent of the equipment being used to erase Joel's memory. A moment later (still 7) a distraught Joel grabs a pole that suddenly disappears from his grasp, an action that is accompanied by a similar buzz. Then, when Joel turns around, he looks behind Clementine, and in the general direction of his gaze, the printing on a sign disappears in a right-to-left wipe (the disappearing sign is between Joel and Clementine, and in still 8 only the first letter of the printing remains). This disappearance is accompanied by a brief musical drum effect. So, in contrast to the first scene, the discontinuities all occur within shots and they do not immediately produce a strong conscious experience. In fact, many viewers miss at least the initial changes (we have shown this in the classroom and most students see only 1 of the discontinuities, if that). As the changes and other events become more salient, they are accompanied by sound cues, and it appears likely that they slowly cross a threshold of awareness that combines with the emotional tone of the scene until viewers finally realize that the scene depicts Joel's memory.

These scenes depict Munsterberg's concept of 'mental play' in ways that only cinema can achieve, and they may have several implications for our understanding of visual cognition. First, the scenes incorporate variations in perceptual salience that can be linked with perceptual experience in interesting ways. Some of the discontinuities produce produce a form of reflective meta-awareness[111] that leads the audience to explicitly consider the link between discontinuity and the memory-erasing procedure, and to understand the sequence as a depiction of imagination rather than real-world events. This idea is familiar to researchers in theory of mind who have proposed that automatic on-line tracking of the attitude associated with an event (e.g., pretending vs believing) provides a crucial interpretive frame for event perception.[113] Similarly, it would be interesting to test whether the cinematic reality of events (as depictions of the characters' real surroundings, vs. their internal mentalizing) reflects a similarly prevasive on-line framework that can strongly influence the interpretation of basic perceptual discontinuities. In addition, it is possible that changes in valence or attitude can be a strong signal that a new event has begun, or that the current event must be reinterpreted[35] and it appears reasonable to consider the possibility that these attitudes are part of visual event representations.

This example also raises interesting questions about the operations that occur in working memory to support event perception. As reviewed above, spatial discontinuities induced by 180 violations induce comparisons in working memory, but the examples in these scenes suggest that the phenomenon could be broader. For example, could a novel multimodal interaction such as the sound-accompanied disappearances in the argument scene afford a pre-to-postchange working memory comparison? To what degree would this comparison be affected by the structural similarity (e.g., matching sound envelopes visual event onsets and offsets) and conceptual similarity between the sound and the visual event? This appears like it might be an interesting open question because on the one hand, research does suggest that some structural links facilitate multimodal integration[114] and event segmentation[115] but on the other hand, research suggests that even affectively powerful and well-matched multimodal stimuli (such as a person scraping a chalkboard with their finger nails) do not necessarily reach awareness.[50]

## CONCLUSION

In introducing this topic, we argued that research exploring the cognitive science of cinema has the potential not only to help understand cinema, but also to change our understanding of visual cognition itself. So, having reviewed this work, how might one characterize these changes? Perhaps the broadest change might be to make clear how, exactly, meaning is related to perception. In isolating the convergence between cinema and real world perception, it becomes clear that both forms of perception can be effective because literal, detailed perceptual codings may be used for identification of objects and movements, but they are not used to integrate events across time. Viewers are incredibly adept at rapidly extracting information from their retinal images when they need it, but they are not locked into the continuing perceptual stream. If they were, cinema would not work because it does not faithfully reproduce this stream, but by the same token, real-world perception would not work because information about the ongoing events that surround us is similarly incomplete. The roots of this argument can be seen in many places, but we believe that research on cinema can provide clear counter-examples to hypotheses about perception that overgeneralize from laboratory tasks that require constant, intensive visual processing and representation.

## FURTHER READINGS

Kraft RN, Cantor P, Gottdiener C. The coherence of visual narratives. *Commun Res* 1991, 18:601–616.

Bordwell D, Carroll N, eds. *Post-Theory: Reconstructing Film Studies*. Madison, WI: University of Wisconsin Press; 1996.

Murch W. *In the Blink of an Eye*. Rev. 2nd ed. Los Angeles, CA: Silman-James Press; 2001.

Shipley TF, Zacks JM. *Understanding Events: From Perception to Action*. Oxford, UK: Oxford University Press; 2008.

## REFERENCES

1. Slide A. *The Encyclopedia of Vaudeville*. Jackson, MS: University Press of Mississippi; 2012.

2. Levin DT, Simons DJ. Fragmentation and continuity in motion pictures and the real world. *Media Psychol* 2000, 2:357–380.

3. Smith TJ, Levin DT, Cutting JE. A window on reality: perceiving edited moving pictures. *Curr Dir Psychol Sci* 2012, 21:107–113.

4. Zacks JM. *Flicker: Your Brain in Movies*. Oxford: Oxford University Press; 2015.

5. Münsterberg H. *The Photoplay: A Psychological Study*. New York: D. Appleton and Company; 1916.

6. Münsterberg H. Why we go to the movies. In: Langdale A, ed. *Hugo Münsterberg in Film: The Photoplay—A Psychological Study and Other Writings*. New York: Taylor and Francis Group; 1915/2001, 171–182.

7. Bordwell D, Carroll N, eds. *Post-Theory: Reconstructing Film Studies*. Madison, WI: University of Wisconsin Press; 1996.

8. Hochberg J. Representation of motion and space in video and cinematic displays. In: Boff KR, Kaufman L, Thomas JP, eds. *Handbook of Perception and Performance*, vol. 1. New York: Wiley Interscience; 1986, 22-1–22-64.

9. Kraft RN. Rules and strategies of visual narratives. *Percept Mot Skills* 1987, 64:3–14.

10. Dmytryk E. *On Film Editing*. London: Focal Press; 1984.

11. Kuleshov L. *Lev Kuleshov: Fifty Years in Film*. Moscow: Raguda Publishers; 1987.

12. Pudovkin VI. *Film Technique*. New York: Random House; 1929/1970.

13. Eisenstein S. A dialectic approach to film form. In: Leyda J, ed. *Film Form: Essays in Film Theory*. New York, NY: Harcourt, Brace, & Jovanovich; 1949, 45–63.

14. Newtson D, Engquist G. The perceptual organization of ongoing behavior. *J Exp Soc Psychol* 1976, 12:436–450.

15. Butterworth G, Jarrett N. What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *Br J Dev Psychol* 1991, 9:55–72.

16. Hood BM, Willen JD, Driver J. Adult eyes trigger shifts of visual attention in human infants. *Psychol Sci* 1998, 9:131–134.

17. Moll H, Tomasello M. 12- and 18-month-old infants follow gaze to spaces behind barriers. *Dev Sci* 2004, 7:F1–F9.

18. Woodward AL. Infants' understanding of the actions involved in joint attention. In: Kail RV, ed. *Advances in Child Development and Behavior*. Oxford: Elsevier; 2005, 229–262.

19. Samson D, Apperly IA, Braithwaite JJ, Andrews BJ, Bodley Scott SE. Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform* 2010, 36:1255.

20. Baker LJ, Levin DT, Saylor MM. The extent of default visual perspective taking in complex layouts. *J Exp Psychol Hum Percept Perform* 2016, 42:508–516.

21. Meltzoff AN. Origins of theory of mind, cognition and communication. *J Commun Disord* 1999, 32:251–269.

22. Apperly IA, Butterfill SA. Do humans have two systems to track beliefs and belief-like states? *Psychol Rev* 2009, 116:953.

23. Leslie AM, Friedman O, German TP. Core mechanisms in 'theory of mind'. *Trends Cogn Sci* 2004, 8:528–533.

24. Lin S, Keysar B, Epley N. Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *J Exp Soc Psychol* 2010, 46:551–556.

25. Zunshine L. Why Jane Austen was different, and why we may need cognitive science to see it. *Style* 2007, 41:275–299.

26. Levin DT, Hymel AM, Baker L. Belief, desire, action, and other stuff: theory of mind in movies. In: Shimamura A, ed. *Psychocinematics: Exploring Cognition at the Movies*. Oxford: Oxford University Press; 2013, 244–266.

27. Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a 'theory of mind'? *Cognition* 1985, 21:37–46.

28. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. *Science* 2004, 303:1634–1640.

29. Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M. Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res* 2009, 2:220–231.

30. Gerrig RJ, Bernardo AB. Readers as problem-solvers in the experience of suspense. *Poetics* 1994, 22:459–472.

31. Hutson J, Smith T, Magliano J, Heidebrecht G, Hinkel T, Tang JL, Loschky L. Eye movements while watching narrative film: a dissociation of eye movements and comprehension. *J Vis* 2015, 15:116.

32. Bezdek MA, Gerrig RJ, Wenzel WG, Shin J, Revill KP, Schumacher EH. Neural evidence that suspense narrows attentional focus. *Neuroscience* 2015, 303:338–345.

33. Visch VT, Tan ES, Molenaar D. The emotional and cognitive effect of immersion in film viewing. *Cognit Emot* 2010, 24:1439–1445.

34. Carroll JM, Bever TG. Segmentation in cinema perception. *Science* 1976, 191:1053–1055.

35. Radvansky GA, Zacks JM. Event perception. *WIREs Cogn Sci* 2011, 2:608–620.

36. Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR. Event perception: a mind-brain perspective. *Psychol Bull* 2007, 133:273.

37. Zacks JM, Kurby CA, Eisenberg ML, Haroutunian N. Prediction error associated with the perceptual segmentation of naturalistic events. *J Cogn Neurosci* 2011, 23:4057–4066.

38. Huff M, Meitz TG, Papenmeier F. Changes in situation models modulate processes of event perception in audiovisual narratives. *J Exp Psychol Learn Mem Cogn* 2014, 40:1377.

39. McNamara DS, Magliano J. Toward a comprehensive model of comprehension. *Psychol Learn Motiv* 2009, 51:297–384.

40. Magliano JP, Dijkstra K, Zwann RA. Generating predictive inferences whil viewing a movie. *Discourse Process* 1996, 22:199–224.

41. Strickland B, Keil F. Event completion: event based inferences distort memory in a matter of seconds. *Cognition* 2011, 121:409–415.

42. Magliano JP, Zacks JM. The impact of continuity editing in narrative film on event segmentation. *Cognit Sci* 2011, 35:1489–1517.

43. Zacks JM, Speer NK, Swallow KM, Maley CJ. The brain's cutting-room floor: segmentation of narrative cinema. *Front Hum Neurosci* 2010, 4:168.

44. Thompson K. *Storytelling in the New Hollywood*. Cambridge, MA: Harvard University Press; 1999.

45. Cutting JE, Brunick KL, DeLong JE. How act structure sculpts shot lengths and shot transitions in Hollywood film. *Projections* 2011, 5:1–16.

46. Chen J, Honey CJ, Simony E, Arcaro MJ, Norman KA, Hasson U. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cereb Cortex* 2015, 155:1–14.

47. Hobbs R, Frost R, Davis A, Stauffer J. How first-time viewers comprehend editing conventions. *J Commun* 1988, 38:50–60.

48. Schwan S, Ildirar S. Watching film for the first time how adult viewers interpret perceptual discontinuities in film. *Psychol Sci* 2010, 21:107–113.

49. Ildirar S, Schwan S. First-time viewers' comprehension of films: bridging shot transitions. *Br J Psychol* 2015, 106:133–151.

50. Troseth GL, DeLoache JS. The medium can obscure the message: young children's understanding of video. *Child Dev* 1998, 69:950–965.

51. Kampe KK, Frith CD, Frith U. "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *J Neurosci* 2003, 23:5258–5263.

52. Wollen KA, Weber A, Lowry DH. Bizarreness versus interaction of mental images as determinants of learning. *Cogn Psychol* 1972, 3:518–523.

53. Pedone R, Hummel JE, Holyoak KJ. The use of diagrams in analogical problem solving. *Mem Cognit* 2001, 29:214–221.

54. Hegarty M, Kozhevnikov M. Types of visual–spatial representations and mathematical problem solving. *J Educ Psychol* 1999, 91:684.

55. Braver TS. The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn Sci* 2012, 16:106–113.

56. Reisz K, Millar G. *The Technique of Film Editing*. New York: Hastings House; 1971.

57. Smith TJ, Henderson JM. Edit blindness: the relationship between attention and global change blindness in dynamic scenes. *J Eye Movement Res* 2008, 2:1–17.

58. Levin DT, Varakin DA. No pause for a brief disruption: failures of visual awareness during ongoing events. *Conscious Cogn* 2004, 13:363–372.

59. Jonides J, Yantis S. Uniqueness of abrupt visual onset in capturing attention. *Percept Psychophys* 1988, 43:346–354.

60. Folk CL, Remington R. Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *J Exp Psychol Hum Percept Perform* 1998, 24:847.

61. Simons DJ. Attentional capture and inattentional blindness. *Trends Cogn Sci* 2000, 4:147–155.

62. Wyble B, Folk C, Potter MC. Contingent attentional capture by conceptually relevant images. *J Exp Psychol Hum Percept Perform* 2013, 39:861.

63. Haber RN. The impending demise of the icon: a critique of the concept of iconic storage in visual information processing. *Behav Brain Sci* 1983, 6:1–54.

64. Levin DT, Simons DJ. Failure to detect changes to attended objects in motion pictures. *Psychon Bull Rev* 1997, 4:501–506.

65. Simons DJ, Rensink RA. Change blindness: past, present, and future. *Trends Cogn Sci* 2005, 9:16–20.

66. Levin DT, Baker LJ. Change blindness and inattentional blindness: from theory to applications. In: Fawcett J, Risko EF, Kingstone A, eds. *The Handbook of Attention*. Cambridge, MA: MIT Press; 2015.

67. Mack A, Rock I. *Inattentional Blindness*, vol. 33. Cambridge, MA: MIT Press; 1998.

68. Simons DJ, Chabris CF. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 1999, 28:1059–1074.

69. Wayand JF, Levin DT, Varakin DA. Inattentional blindness for a noxious multimodal stimulus. *Am J Psychol* 2005, 118:339–352.

70. Shapiro KL, Arnell KA, Raymond JE. The attentional blink: a view on attention and a glimpse on consciousness. *Trends Cogn Sci* 1997, 1:291–296.

71. Kanwisher N. Repetition blindness: type recognition without token individuation. *Cognition* 1987, 27:117–143.

72. O'Regan JK, Rensink RA, Clark JJ. Change-blindness as a result of 'mudsplashes'. *Nature* 1999, 398:34.

73. Schoeberl T, Fuchs I, Theeuwes J, Ansorge U. Stimulus-driven attentional capture by subliminal onset cues. *Atten Percept Psychophys* 2015, 77:737–748.

74. Simons DJ, Levin DT. Failure to detect changes to people during a real-world interaction. *Psychon Bull Rev* 1998, 5:644–649.

75. Levin DT, Simons DJ, Angelone BL, Chabris CF. Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *Br J Psychol* 2002, 93:289–302.

76. Simons DJ, Chabris CF, Schnur TT, Levin DT. Evidence for preserved representations in change blindness. *Conscious Cogn* 2002, 11:78–97.

77. Chabris CF, Weinberger A, Fontaine M, Simons DJ. You do not talk about Fight Club if you do not notice Fight Club: inattentional blindness for a simulated real-world assault. *i-Perception* 2011, 2:150–153.

78. Raymond JE, Shapiro KL, Arnell KM. Temporary suppression of visual processing in an RSVP task : an attentional blink? *J Exp Psychol Hum Percept Perform* 1992, 18:849–860.

79. Rensink RA, O'Regan JK, Clark JJ. To see or not to see: the need for attention to perceive changes in scenes. *Psychol Sci* 1997, 8:368–373.

80. Hollingworth A, Henderson JM. Semantic informativeness mediates the detection of changes in natural scenes. *Vis Cogn* 2000, 7:213–235.

81. Werner S, Thies B. Is "change blindness" attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Vis Cogn* 2000, 7:163–173.

82. Brockmole JR, Hambrick DZ, Windisch DJ, Henderson JM. The role of meaning in contextual cueing: evidence from chess expertise. *Q J Exp Psychol* 2008, 61:1886–1896.

83. Frith U, Robison JE. Perceiving the language of films. *Perception* 1975, 4:97–103.

84. Huff M, Schwan S. Do not cross the line: heuristic spatial updating in dynamic scenes. *Psychon Bull Rev* 2012, 19:1065–1072.

85. Berliner T, Cohen DJ. The illusion of continuity: active perception and the classical editing system. *J Film Video* 2011, 63:44–63.

86. Baker LJ, Levin DT. The role of spatial triggers in event updating. *Cognition* 2015, 136:14–29.

87. Levin DT, Wang C. Spatial representation in film. *Projections* 2009, 3:24–52.

88. Hermer L, Spelke ES. A geometric process for spatial reorientation in young children. *Nature* 1994, 370:57–59.

89. Ratliff KR, Newcombe NS. Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cogn Psychol* 2008, 56:142–163.

90. Mandler JM, Seegmiller D, Day J. On the coding of spatial information. *Mem Cognit* 1977, 5:10–16.

91. Schulman AI. Recognition memory and the recall of spatial location. *Mem Cognit* 1973, 1:256–260.

92. Shoqeirat MA, Mayes AR. Disproportionate incidental spatial-memory and recall deficits in amnesia. *Neuropsychologia* 1991, 29:749–769.

93. Zwaan RA, Radvansky GA. Situation models in language comprehension and memory. *Psychol Bull* 1998, 123:162–185.

94. Zwaan RA, Van Oostendorp H. Do readers construct spatial representations in naturalistic story comprehension? *Discourse Process* 1993, 16:125–143.

95. Hochberg J, Brooks V. Movies in the Mind's Eye. In: Bordwell D, Carroll N, eds. *Post-Theory: Reconstructing Film Studies*. Madison, WI: University of Wisconsin Press; 1996, 368–387.

96. Hollingworth A. Constructing visual representations of natural scenes: the roles of short-and long-term visual memory. *J Exp Psychol Hum Percept Perform* 2004, 30:519–537.

97. Magliano JP, Miller J, Zwaan RA. Indexing space and time in film understanding. *Appl Cogn Psychol* 2001, 15:533–545.

98. Levin DT. Spatial representations of the sets of familiar and unfamiliar television programs. *Media Psychol* 2010, 13:54–76.

99. Epstein RA. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 2008, 12:388–396.

100. Magliano JP, Larson AM, Higgs K, Loschky LC. The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Mem Cognit* 2016, 44:207–219.

101. Huff M, Papenmeier F, Zacks JM. Visual target detection is impaired at event boundaries. *Vis Cogn* 2012, 20:848–864.

102. Shimamura AP, Cohn-Sheehy BI, Shimamura TA. Perceiving movement across film edits: a psychocinematic analysis. *Psychol Aesthet Creat Arts* 2014, 8:77–80.

103. Claus B, Kelter S. Comprehending narratives containing flashbacks: evidence for temporally organized representations. *J Exp Psychol Learn Mem Cogn* 2006, 32:1031–1044.

104. Raisig S, Welke T, Hagendorf H, van der Meer E. I spy with my little eye: detection of temporal violations in event sequences and the pupillary response. *Int J Psychophysiol* 2010, 76:1–8.

105. Hymel A, Levin DT, Baker LJ. Default processing of event sequences. *J Exp Psychol Hum Percept Perform* 2016, 42:235–246.

106. Hagmann CE, Cohn N. The pieces fit: constituent structure and global coherence of visual narrative in RSVP. *Acta Psychol (Amst)* 2016, 164:157–164.

107. Sirigu A, Zalla T, Pillon B, Grafman J, Agid Y, Dubois B. Encoding of sequence and boundaries of scripts following prefrontal lesions. *Cortex* 1996, 32:297–310.

108. Smith TJ. The attentional theory of cinematic continuity. *Projections* 2012, 6:1–27.

109. Levin DT, Saylor MM. Shining spotlights, zooming lenses, grabbing hands, and pecking chickens: the ebb and flow of attention during events. In: Shipley T, Zacks J, eds. *Understanding Events: From Perception to Action*. New York: Oxford University Press; 2008, 522–554.

110. Chun MM, Golomb JD, Turk-Browne NB. A taxonomy of external and internal attention. *Annu Rev Psychol* 2011, 62:73–101.

111. Schooler JW, Smallwood J, Christoff K, Handy TC, Reichle ED, Sayette MA. Meta-awareness, perceptual decoupling and the wandering mind. *Trends Cogn Sci* 2011, 15:319–326.

112. Villena-González M, López V, Rodríguez E. Orienting attention to visual or verbal/auditory imagery differentially impairs the processing of visual stimuli. *Neuroimage* 2016, 132:71–78.

113. Leslie AM. Pretending and believing: issues in the theory of ToMM. *Cognition* 1994, 50:211–238.

114. Vroomen J, Gelder BD. Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform* 2000, 26:1583–1590.

115. Saylor MM, Baldwin DA, Baird JA, LaBounty J. Infants' on-line segmentation of dynamic human action. *J Cogn Dev* 2007, 8:113–128.