

Exploring Inner Worlds: where cognitive psychology may take us

A dialogue between Tim J. Smith and Ian Christie

Tim Smith researches visual cognition through a variety of techniques that capture what subjects do when they watch naturalistic visual scenes. Although much research into scene perception uses static visual scenes, he is interested in how we process dynamic visual scenes, including feature films. Using techniques such as tracking the eye movements of viewers watching film sequences in combination with behavioural probes during and after viewing, he investigates the cognitive processes occurring during film viewing. His research has contributed to a computational model of fixation durations in scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010), and he has proposed an Attentional Theory of Cinematic Continuity (Smith, 2012), which has attracted the interest of 'cognitivist' film scholars, leading to his involvement in a number of presentations and projects that use empirical testing and modelling to explain what happens when we watch film sequences.

I too am interested in the long history of attempts to understand the perceptual and cognitive processes involved in film viewing, which stretch from pioneering work by Munsterberg and Rank in the 1910s, through the renewed interest of Russian psychologists in the 1920s, following Pavlov's classic studies in conditioning, and later of the Gestalt school, up to present-day work such as that of James Cutting (starting from the 'psychophysics' tradition) and Uri Hasson ('neurocinematics'). In this exchange, written as a dialogue over several months, I invite Tim Smith to summarise some of his own work and to speculate on what value it may have for non-scientists interested in understanding how and why we perceive film as we do – and how this may be changing. Ian Christie

IC: You've said that the Hollywood style of filmmaking, which permeates a wide range of visual media, has evolved formal conventions that are *compatible with the natural dynamics of attention and humans' assumptions about continuity of space, time, and action*. It's the word 'natural' that interests me in relation to where we are now in the history of moving-image entertainment - especially since we know that 'Hollywood style' has changed considerably, and continues to change. Are you starting from the assumption that 'Hollywood' represents a form of spectacle optimised to match a mass audience's interests and abilities?

TJS: Yes, my empirical and theoretical investigations into the editing conventions that are a part of the Hollywood style seem to suggest that they create an audiovisual spectacle that is in tune with the existing constraints and expectations of viewer cognition. For example, the convention of using consistent screen direction of a character's movement across a series of shots to create an impression of continuity of their motion matches the way we would follow the same action if we viewed it in the real-world from a static viewpoint. Imagine how your eyes and head would move if you were watching a person walk in front of you in the real-world. You first follow their movement by making a series of pursuit movements with our eyes then make a head rotation along the direction of their motion as the character moves out of our comfortable field of view. Our eyes then have to compensate for the head movement by making quick saccadic eye movements in the opposite direction. This pattern of eye movements is very similar to how we would watch the same action depicted across a series of shots. The only difference is that the cut replaces the head movement.

I argue that these conventions are 'natural' as they accommodate the way we would normally attend to such scenes in the real-world. Clearly there are differences between a scene

presented in the real-world and the same scene presented as an edited sequence but these differences are minimised so that viewers can perceive the depicted actions with minimal learning. This idea isn't new. In fact, one of the first film theorists, Hugo Munsterberg (1916) believed that the "photoplay" had developed techniques such as the close-up, shot/reverse-shot sequence and the analytical breakdown of the scene so rapidly because it externalized the audience's inner world. Close-ups were the cinematic equivalent of focussed attention. Flash-backs, flash-forwards and elastic representations of duration replicate the way our memory influences our perception of time. This cognitive compatibility of the Hollywood style has since been acknowledged by other theorists such as J.J Gibson, Julian Hochberg and Virginia Brooks and the Cognitive Film Theorists including Joseph Anderson. Anderson laid out the framework for an ecological view of film perception in his book *The Reality of Illusion* (1996).

As for the ever changing form of the Hollywood Style, some people have seen this as evidence against the ecological view of film perception. How can a style that is meant to represent the cognitively 'optimal' way of representing a scene change so radically from the long-take aesthetic of a director such as Billy Wilder in the 1950s and 60s to today's rapidly edited and highly mobile blockbusters such as those directed by Michael Bay and Roland Emmerich? Even respected directors who have been in the business since the 60s, such as Ridley Scott have adapted their style over time and now incorporate greater pace and mobility than they would in their earlier films. Such development of the Hollywood style can be explained cognitively in two ways. The Hollywood Style may have always been evolving towards the cognitive optimum even during periods of apparent stability. Or audiences are changing and film adapts to the new demands of audiences.

Before we get to changing style, you mention Munsterberg believing that the films of the 1910s, when he was writing, 'externalized the audience's inner world'. The way he put it was that films 'adjusted events to the forms of the inner world, namely, attention, memory, imagination, and emotion'. What's interesting about this, both then and now, is the idea that something like what we see constructed on screen is *already present* as a kind of 'inner movie'. That film replicates in some way how we already imagine space, time and events – like a visual form of what another pioneer psychologist, Lev Vygotsky, called 'inner speech'.

When we begin referencing 'inner' worlds, whether they be speech or space-based we enter into unstable territory. There is no way of knowing what form these mental representations of external stimuli take and they are malleable over time and experience. If we expose a person to a candidate representation such as a film and ask them whether the candidate matches how they perceived a space, the candidate infects their internal representation, creating a hybrid of both. A film can never be an ideal representation of one person's experience of a space as the internal representation is never stable. For example, in the opening 'Choose Life' sequence of *Trainspotting* (Danny Boyle, 1996), Renton and Spud run east down Princes Street in Edinburgh, chased by security guards from a shop they have just robbed. In one shot they are in the middle of Princes Street and in the next shot they turn down a flight of stairs on to Calton Road just as Renton is run over by a car. The match-on-action between the shots implies that the two shots are spatially contiguous but having lived in Edinburgh for thirteen years, I know that it would have taken Renton several minutes to run between the two locations. My mental representation of the space negates the fantastical space created by Boyle, potentially bringing me out of the narrative momentarily as I notice the discrepancy. However, I first viewed *Trainspotting* before I moved to Edinburgh and I remember watching the opening sequence and perceiving the two locations as contiguous. Danny Boyle has created a representation of space that is valid for his narrative purpose and as an audience of this filmic representation I can perceive it both as a valid narrative space and an invalid representation of the actual geography

of Edinburgh. Sometimes it can feel as if I hold these mutually exclusive representations of the same space active in my mind at the same time. In reality I am probably rapidly switching between the two percepts as in the bistable figure of the vase/faces below. The image can be perceived either as two black faces looking at each other or as a white vase but never both at the same time. This is an example of how perception is not just about seeing what is there in front of us. It is as much about formulating hypotheses about what we expect to be there based on prior experience. In the *Trainspotting* scene, if I attend to the space depicted in the scene I will perceive a spatial discontinuity. But if I attend to Renton's movement and the narrative the discontinuity will be less salient. Now the outstanding question is whether attending to the narrative means that I fail to represent the space or whether I construct a mental representation of the space but do not notice the discontinuity because its importance has been downgraded by lack of attention.

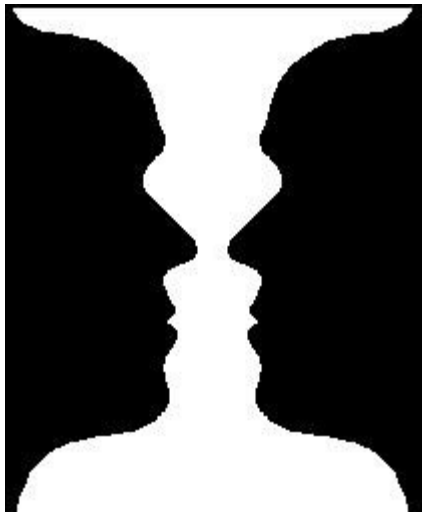


Figure 1: A bistable figure. We can either perceive two black faces looking at each other or a white vase but we cannot perceive both at the same time.

One question I wish to raise in response to the Munsterberg quote is the issue of who's internal representation a film is meant to be? If a film is a medium of communication between the director and the viewer then the film has to be both a representation of how the director perceives the scene and how they want the viewer to perceive the scene. As most fiction film places the emphasis on creating a narrative rather than documenting a space or event, the director must prioritize the viewer's percept. The director's experience of a film will be confounded by their knowledge of all aspects of the film's production, e.g. how the script was adapted, how the scene was staged, how many takes were shot, what the weather was like on the day of the shoot and the taste of the stale bagel he ate whilst waiting for the lights to be hung. These memories are part of the director's internal representation of a film scene and must be ignored if the narratively significant elements of the scene are to be conveyed to a viewer. Munsterberg was right to observe that film "adjusted events to the forms of the inner world" as a director has to ignore their own percepts and focus instead on simplifying and anticipating how a viewer will construct their own inner world of a film. The ability to invoke intended internal states in a viewer through minimal external cues is one of the magical qualities of film, just as in literature, music, and art.

IC: Magical, yes certainly, sometimes. But I think I'd also want to describe it as 'transformative'. Film transforms its material, whether this is 'raw', like the streets of Edinburgh, or a studio set contrived to look like Edinburgh, or indeed like Neverland - a place we've never seen, but which

the film's makers want us to believe is 'real', at least while watching the film. But you're right to question Munsterberg's remark about internal, or at least personal representations and the objective public film. The director, we might say, works to optimize what has been created for what s/he assume will be the largest number of intended spectators. And I know some editors who refuse to attend the shooting so that they won't be influenced by what they see there, but want to consider the footage 'as it is'. But is there scope in your methodology of recording viewer response to challenge the traditional idea in film studies that 'we' all see the 'same' film? Or does individual viewer response vary enough to support the idea of a spectrum of perception, and hence response. Do perception and cognition studies create potentially a new paradigm for spectatorship - or just finesse what we already think we do/know?

TJS: I didn't know that editors often refuse to witness the shoot. That makes perfect sense as our memories are constantly being updated by new perceptual experiences and our perceptual experiences are always cast in the shadow of our memories. The editor has to have fresh eyes on the screen content if they are to understand how the audience will see the finished sequence. The degree of insight good film editors seem to have into viewer cognition is astounding. Often this insight is tacit and the editor does not or cannot express it. Instead, like a lot of craftspeople they "feel" when a cut is right by reflecting on their own perception of a scene. Occasionally, there are editors who can express their insights and describe their techniques. Walter Murch, Karel Reisz, and Edward Dmytryk are the finest examples of this rare breed. In several books on the topic of film editing and direction they describe techniques they have used to create fantastical spaces out of minimal film material. As you say, editing is often 'transformative', creating the impression of a coherent and plausible space from minimal viewpoints probably shot in different locations, at different times and using only minimal sets and props. Walter Murch has described the discovery that scenes could be created out of an edited sequence of discontinuous viewpoints "the cinematic equivalent of the discovery of flight" (Murch, 2001).

This constructive nature of film perception was first noted and studied by Lev Kuleshov and Vsevolod Pudovkin in 1920's Russia (Pudovkin, 1929). Across a series of "pseudo-experiments" - I say pseudo as they were filmmakers, not Psychologists and their experiments didn't contain the necessary controls for us to draw strong conclusions - Kuleshov and Pudovkin created film sequences in which they manipulated the way in which scenes were presented across edits. The most classic example is known as the 'Kuleshov' or 'Mosjukhin Experiment' after the actor depicted in the footage. In this experiment, a close-up shot of the actor Mosjukhin's face was juxtaposed with a close-up of either a bowl of soup, a young girl or a dead woman. Even though the actor's face displayed no recognisable emotion when viewed in isolation viewers described the face as displaying hunger and longing when followed by the soup. When followed by the young girl they saw a happy, light smile on Mosjukhin's face but saw sorrow and loss on the same face when followed by the dead woman. Kuleshov and Pudovkin had demonstrated the constructive power of film: the juxtaposition of shots can create a meaning not present in either. This idea was taken to its logical extreme by their peer Sergei Eisenstein who used the collision of two seemingly unrelated shots to encourage the viewer to interpret one in relation to the other. For example, In Eisenstein's *October* (1928) a shot of a mechanical golden peacock, not belong to the storyworld, placed next to a shot of a man leads the viewer to conclude that the man is vain. Eisenstein's technique of *dialectical montage* was very different to Kuleshov and Pudovkin's *constructive montage*, in which the juxtaposition of shots containing minimal cues to their continuity, such as Mosjukhin looking out of frame or an actor walking from one shot to another gives the impression of a continuous scene even if, in reality the shots were not filmed in the same location. Surprisingly, even though this phenomenon was first investigated over 80 years ago psychologists have only recently begun examining this phenomenon and using it to

reflect on theories about how we perceive real-world spaces. Along with Dan Levin and James Cutting I have recently reviewed this literature in an article for *Current Directions in Psychological Science* (2012).

I also want to pick up on your question about “new tools for studying spectatorship”. I absolutely agree that the experimental techniques and measurement devices from the cognitive sciences can advance our understanding of how we watch movies. There has been a lot of recent evidence for a consistency in how multiple viewers watch movies (Smith, in press) and even how our brains process movies (Hasson et al, 2002). I have shown across a few studies that if we use an eyetracker to monitor the gaze location of multiple viewers when watching most film sequences, the location of our gaze will demonstrate a remarkable degree of consistency. I refer to this spontaneous clustering of gaze as *attentional synchrony* (Mital, Smith, et al, 2011). This is a characteristic of film spectatorship we never would have known about if it weren't for eyetracking. Filmmakers and theorists may have hypothesised such consistency but this would have remained a hypothesis without a method for testing it.

By describing the average behaviour of a viewer during a film sequence we also have a way of identifying idiosyncratic viewing behaviours. Continuing with eyetracking as our example (but this idea equally applies to brain imaging or recording biophysiological responses such as heart rate), recording the gaze behaviour of multiple viewers gives us a distribution of gaze points for each frame of a movie. The mean gaze point of this distribution will be the position that best represents the majority of viewers. For example, when looking at any medium close-up of an actor the mean gaze point will generally be centred on their face. The further a screen location is away from the face the likelihood of a viewer looking in that position gets less and less. But, imagine that the medium close-up came from a horror movie and a subset of viewers were seasoned horror fans. The mean gaze point of the whole set of viewers might still be located over the protagonist's face but the distribution of gaze points may also show that our horror buffs are statistical outliers in that their gaze is often focussed over the protagonist's shoulder as they await the sudden appearance of the psycho killer in the background. By examining the distribution of gaze we can identify this cluster of viewers who share a different mode of viewing. The same technique could be used to identify expertise in viewing, expectation of narrative content, gender differences, or the impact of prior experience with the movie or source novel on how we watch movies. In combination with behavioural measures such as memory tests for content, measurements of emotional responses or brain activity, we can triangulate the cinematic experience and directly test hypotheses about spectatorship.

IC: I wonder if this kind of investigation of ‘deviations’ from normal ‘attentional synchrony’ could ultimately shed some light on what happens when we view older films today? At the time that the early Soviet pioneers were doing their practical experiments and speculating about how perception could be stimulated, or even retrained in terms of Pavlovian conditioning, they were working essentially in the present. (Although it's interesting that the famous ‘Mosjoukine effect’ was tested with footage of a pre-revolutionary star, who had gone into exile after the revolution: I suspect they might have used used him because his face was still very familiar to Russian viewers). But today we watch films from the whole history of cinema, including a lot from the ‘silent’ period, which arguably means that some of us at least have had to become ‘expert’ in older modes of perception. Visually, this is rather like the art historian Michael Baxendall's idea of a ‘period eye’ being required to grasp the art of the past; or indeed what happens when we listen to early pre-classical music and accept its harmonic structures as very different - more limited, if you like - than what came later, and has shaped our contemporary musical expectations.

But I'd like to move this on to the issue of contemporary viewing, and the claim that's often

made that our perceptual norms for 'screen entertainment' or content are, for instance, being accelerated (faster cutting) and perhaps more fundamentally re-organised in spatial terms by new fashions in 'ungrammatical' shot-linkage. Do you see any empirical evidence for new regimes of spectatorship being produced by new media styles?

TJS: Film form is definitely changing and viewer expectations of how films should 'stimulate' them also seem to be changing. Film statistician, Barry Salt has made a career of quantifying the change of film form over time. His statistical analysis of such features as shot length, shot size, and transition type (e.g. fades, dissolves, etc) has spawned an entire approach to film analysis known as *cinematics*. Several film theorists have followed in Salt's footsteps, most influential of whom is James Cutting, a perceptual psychologist. Cutting and his team have analysed the formal features of 160 Hollywood films from 1935 to 2010 and identified a significant decrease in average shot length from 15 seconds to 3.5 seconds (Cutting, et al, 2011). They have also noted a significant increase in the amount of motion from frame to frame and a significant decrease in the average brightness of movies. Being a psychologist, Cutting explains these changes in relation to an attempt by film makers to increase control over how viewers watch films. If we assume that cuts and motion force the reorientation of the viewer's attention, and that darker images give viewers fewer options of where to look then all of these changes should lead to greater control over viewer attention.

However, such control over viewer attention is not new. In fact, I argue in my *Attentional Theory of Cinematic Continuity* (AToCC; Smith, 2012) that coordinating viewer attention is a critical component of the classical Hollywood style of editing. As I have already mentioned, continuity edits resemble the way we would attend to a similar scene in the real-world, with cuts replacing the natural shifts of attention from one part of the scene to another. Classic continuity editing has always used visual cues such as onsets of motion to capture and hold attention (e.g. in a Match Action cut) and shot composition and *mise en scene* guide attention to the narratively relevant parts of a scene. The recent changes in film form documented by the likes of Cutting are, as David Bordwell has described them, an *intensification* of the classic continuity style (Bordwell, 2006). Bordwell argues that the current mainstream Hollywood style of filmmaking is characterised by four main stylistic changes: rapid editing, bipolar extremes of lenses (i.e. rapid changes in depth of field), reliance on close-shots, and wide-ranging camera movements. Any longterm film viewer can confirm these changes. For example, if we compare the fight sequences from a 60s or 70s Bond film such as *Moonraker* (1979) to fight scenes in *Casino Royale* (2006) or *Quantum of Solace* (2008), the earlier sequences seem slow and hesitant. Mostly shot in long or medium shots with shots lasting several seconds, they allow us to see all of the action from a distance. Whilst providing a clear sense of space and the relationship between characters these earlier scenes lack the intensity and brutality seen in the more recent Bond movies. Inspired by the Bourne series, most recent fight sequences are shot very close with rapidly edited shots taken by a mobile camera. This style is favoured by modern directors as it seems to drive viewer attention, create a heightened state of arousal and disorientation suitable for the on-screen action.

In my analysis of viewer gaze behaviour during modern film sequences, such as these from the Bond movies I have shown that intensified continuity leads to a greater degree of coordination between where viewers look on the screen and a tendency to fixate the screen centre without much exploration of the rest of the screen (Smith, in press). This centre bias is necessary as the editing rate and camera movement is so extreme that viewers do not have time to recover from a cut and move their eyes to a new part of the screen before the next cut happens. This results in a strange mismatch between a heightened sense of activity created by the intense editing but an actual passivity in how viewers attend to the film. All viewers look in the same place but they

do so because the director is forcing them to, not because they choose to. Over prolonged periods such overt manipulation of viewer gaze can lead to fatigue and habituation to the normal peak in arousal that would be experienced by harsh cuts. This is why we can eventually grow bored and indifferent to an overly long action sequence (Michael Bay, I'm looking at you!).

Now consider the opposite approach to guiding viewer attention: the slow film. Instead of cutting to reframe the scene, moving the camera to point at an object of interest or racking focus to pull the eyes about in the frame, what about if we just let the shot linger? You may think that the absence of any of these cinematic techniques would lead to less control over where viewers look. But considering that our visual system evolved to deal with a continuous viewpoint on a real-world scene we must have ways to distinguish what is relevant from irrelevant in any scene. The information we use to make this decision is typically characterised as either *exogenous* (in the outside world) or *endogenous* (within us). Exogenous cues are features of the real-world that involuntarily capture our attention such as the sudden onset of motion or the turning on of a light. Attentional capture by such features is relatively automatic and similar across all viewers. Hence why Michael Bay's big explosions will reliably lead to attentional synchrony (at least until we habituate to them). By comparison, endogenous factors refer to how we perceive a scene, how it relates to our memories and what we expect to see. Endogenous factors can vary massively across viewers leading to a lack of attentional synchrony. Films that rely on endogenous control of attention can also be perceived as more effortful. For example, slow films such as *Once Upon a Time in Anatolia* (Ceylan, 2011), or the films of Bela Tarr, often contain long takes in which not much seems to happen. The viewer is free to explore the frame (using endogenous control), interrogating the scene and the actors to find something of interest to look at. Such 'slow films' have recently been the topic of a large debate on-line and in print about whether the critical acclaim often attributed to such films is an example of film critics mistaking boredom for High Art. In his article for the *New York Times* (April 29th, 2011), film critic Dan Kois compared watching slow films to eating his 'cultural vegetables': he didn't enjoy it, but because somebody told him it was good for him he felt he should do it.

Kois's article led to several film critics defending slow film and the enjoyment and sense of reward that can be gained from the effortful viewing of a non-traditional film such as Paul Thomas Anderson's *There Will be Blood* (2007). The sense of accomplishment the critics are experiencing is due to their ability to endogenously control how they attend to the film and what they perceive in it. In part this is due to the fact that they have learnt to watch such films and see more in each shot than is immediately apparent to a naïve viewer. But the sense of satisfaction can also come from the simple pleasure of reveling in the human form in motion and interacting with other bodies. Intensified films, such as the *Transformer* series, bombard us with an indiscipherable series of close-up shots designed to drive the viewer exogenously without engaging higher cognition. Often the human (or robot in this case) form and its relationship to other bodies is lost. In my analysis of how we watch *There Will Be Blood* I have observed spontaneous moments of attentional synchrony in the absence of any editing or cinematography techniques (Smith, 2012). The simple turn of a head, movement of a hand, or onset of speech is enough to reconcile where all viewers look. Through exquisite direction and acting a scene shot in a single long take with minimal action can deeply engage its viewers and make them active participants in the scene.

So yes, slow films can require a level of endogenous control beyond what a lot of film viewers are willing to give but the rewards can outweigh the costs. Today's slow films have to work harder to engross audiences than the same pace of film would in the past, since they are being seen against the background of blockbusters operating purely exogenously. As you said, experienced film viewers can watch a silent movie, a Russian formalist film, or a Hollywood

musical and adopt a 'period eye' that allows them to appreciate the film within the context in which it originally appeared. Our knowledge of film style, film history and the technical limitations of the era allow us to adopt a viewing style that sees through the film to its narrative core and limits our awareness of the formal features that may deter a less experienced viewer. Whether such contemplative films will survive the continued onslaught of information in society and the multiplicity of screens fighting for our limited attention we will have to wait and see. I hope so, for the sake of good cinema and society.

IC: And can the same be said - or expected - for the onslaught of 3D and proposed 'sensory feedback' set-ups like the one proposed by Filmtrip? (f/n on this) My sense is that these are not a problem for the continued health of cinema (or society), in the sense that they extend and multiply the range of perceptual options available to us. But I suppose you could argue that in doing so they distract from the training needed for such specialised perceptual tasks as, say, watching Murnau, Eisenstein, Ozu or Bela Tarr. Can you see any evidence of perceptual - or should this be considered attentional - confusion or overload from your experiments, or indeed from your own experience as a spectator?

TJS: Cinema is in a time of flux and uncertainty. The proliferation of presentation formats (high-def, IMAX, 3D, higher frame rates) seems to be evidence of a medium losing its way and desperately trying to find its feet against the background of the new media cultures of videogames, the internet and mobile technologies. However, I agree with you that this technological panic does not indicate the decline of cinema. Instead I think it will lead to a refinement of the key joys of the cinematic experience, and force the film industry and cinema audiences to focus on these elements instead of trying to compete with other more interactive media. For me, the essence of the cinematic experience is the audience's relinquishing of control to the director under the expectation that the director will craft an audiovisual experience that will entertain, intrigue, challenge or move the audience. A film cannot compete with videogames because we enter into the experience with completely different expectations about what we want to get out of it. I play a first-person shooter because I want to have the thrill of control, of mental and physical challenge, and the visceral response to an environment and situation that would be too dangerous for me to experience in real-life. When I go to the cinema to watch a film depicting a similar war scenario, such as Spielberg's *Saving Private Ryan* (1998) I expect to gain an insight into how it felt to be a soldier landing on Omaha beach, but I do so through empathy, immersion and engagement with the narrative not through physical agency. If cinema attempts to steal videogame audiences by adding agency into the film experience, I think it will be making a crucial mistake and risking the loss of the richer, unique ability of film to involve us emotionally and cognitively in a storyworld without agency.

However, the two technologies you mention, stereoscopic 3D and physiologically responsive film do not attribute conscious agency to the audience and I believe they both, therefore have a potential to advance cinema. Turning first to stereoscopic 3D, the recent mass adoption of 3D by film companies and the push by technology companies for 3D capabilities to be added to all digital screens has meant that there is great potential for cinema to reflect on the assumed limitations of 2D filmmaking and add 3D as a new creative tool. However, it also creates the risk that film companies will go for the quick buck and use cheap thrills that audiences will grow tired of very quickly. As I write this, there is already evidence that the audience for 3D movies is decreasing and there is no longer a guaranteed boost to ticket sales with the addition of 3D. Part of this backlash is probably due to audiences feeling manipulated by higher priced 3D tickets that don't add anything to the film and may even make it more uncomfortable to watch. However, this doesn't have to be the case. Used with a creative purpose and with insight into what 3D is good and bad for, I think some films can benefit from 3D.

There have already been a few examples of wonderful uses of 3D. *Avatar* (James Cameron, 2009), *Hugo* (Martin Scorsese, 2011), *The Cave of Forgotten Dreams* (Werner Herzog, 2011), *Pina* (Wim Wenders, 2011), *Coraline* (Henry Sellick, 2009) were designed from conception as 3D movies, and as such they make exquisite use of the potential for 3D to render spaces and actions in a way that 2D can only imply. Unfortunately, these fine examples of 3D are cast against a mass of poor 3D films or films converted into 3D after filming that simply use 3D as a cheap trick, poking things at the audience instead of creating enacted volumes. I believe that 3D will survive in cinema, but it may stop being the norm for blockbusters and instead be reserved for a small number of films that wish to tackle its complexities and use it in a creative way. Filmmakers who take on this task must be willing to engage with the psychology of their viewers and understand how our eyes deal with the stereoscopic images in order to perceive depth. For example, some modern conventions of film composition such as rapid camera movement and fast cutting don't work in 3D, because our eyes take longer to adjust to new scenes and motion presented at 24 frames per second creates uncomfortable disparities between our two eyes. The limited empirical research that has been conducted on 3D film viewing has clearly shown that these difficulties are evident in viewer eye movements (Hakkinen et al, 2010). But if 3D scenes are composed with due regard for these problems, viewers are much more active in 3D than 2D scenes, exploring more of the frame and not just attending to such focal features as faces. Slowing down cutting and keeping the camera more stable allows viewers to overcome these issues and view 3D more comfortably. In fact, adopting some of the visual narrative techniques of the *slow* filmmakers previously mentioned may serve 3D filmmakers very well. Stereoscopic 3D filmmaking is in a rapid state of evolution, but once such considerations become widely known I believe filmmakers will use it to create cinematic experiences never previously thought possible.

As for pseudo-interactive cinema experiences like Filmtrip, I also believe we will see more of these techniques in the future. Cinema has a long history of trying new techniques to enrich the experience and attract more paying customers. Participatory experiments such as Smell-a-Vision, William Castle's vibrating *The Tingler* (1959) and motion simulation have attempted to increase the audience's feeling of being part of the action. Such techniques attempt to maximize immersion in the storyworld. They seem to be motivated by making film more like Virtual Reality. I inherently disagree with this approach. The intention of film is not to transport you to another world. If it were, multiplexes would be full of wildlife documentaries instead of narrative fictions.

Instead of VR, I think oral storytelling is a better analogy for the kind of cinematic experience proposed by Filmtrip. In oral storytelling, a skilled storyteller weaves a story together from a series of critical elements that can be presented in innumerable ways depending on the particular audience. The skill of the storyteller is to adapt the story to the audience in a way that optimizes the experience for them. For instance, when telling the tale of Little Red Riding hood to a group of children the storyteller might moderate the more horrific elements of the story in favour for the life lesson of not trusting strangers. The same storyteller, telling the same story to a group of adults may rack up the gore and dwell more on the sexual undertones of a young girl venturing forth alone into the world and being enticed into a wolf's bed. The telling of the story depends on the feedback the storyteller gets from the audience. Such feedback is rarely in the form of instructions about what should happen next - as it would be in an experience with more agency such as a videogame - instead there are the gasps, the giggles, the screams and the confused faces of the audience which the storyteller registers and folds into how they tell the story. To date, cinema has been blind to the moment-by-moment comprehension and emotion of the audience. The only chance filmmakers have had to get feedback from audiences has

been through test screenings in which an audience is invited to see early versions of film in production and fill in a questionnaire after the screening.

However, such information is useless for identifying exactly how and when a film lost the audience. *Reactive Cinema*, as I would call it, uses real-time observational data from the audience to decide whether their collective emotional and cognitive state is as desired, and if not to try and modify the audio and visuals in a way that can get the audience back on track. There are many techniques for measuring implicit audience feedback, including heart rate, galvanic skin response, eye tracking, facial expression monitoring, agitation (“edge of seat-ness”) or electroencephalography (EEG). The questions we face right now is knowing how to infer the relevant states from these measurements (or a combination of them), and knowing how best to modify films. I highly commend research groups working in this field, such as Filmtrip, and look forward to seeing where this research takes us.

IC: We may be among the relatively few who do! But I suspect that the current plenitude of viewing experiences that’s available to cinephiles may actually be creating multi-skilled viewers, who can adjust to extreme differences of screen size, not to mention the absence of synch sound (in the burgeoning of ‘silent’ film screenings at specialized festivals like Pordenone and Bologna), and extremes of shot scale and editing rhythm. Much is said about modern spectators having short attention spans, but if we look back at the range of what cine-literate audiences had to deal with between the 1950s and the 1980s, I would argue that present-day audiences are actually being ‘stretched’ more, and numbers of them seem to be developing highly adaptive skills, including those of multi-tasking *while* ‘watching a film’ (as discussed by Roger Odin elsewhere in this collection). What we have long called ‘cinema’ is changing faster than at almost any time in nearly a hundred and twenty years of moving pictures and an even longer period of recorded sound. As you have often said, it’s amazing that these phenomena, which occupy the largest part of our imaginative and cultural worlds, have attracted such a small amount of fundamental research. Fortunately, this seems to be changing, and I look forward to your future work, and that of all the other scientists who are now taking movies seriously.

References

- Anderson, J. (1996). *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Southern. Illinois University Press.
- Bordwell, D. (2006). *The Way Hollywood Tells It: Story and Style in Modern Movies*. Los Angeles, US: University California Press.
- Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, 2, 569-576.
- Häkkinen, J., Kawaid, T., Takataloc, J., Mitsuyad, R. & Nymanc, G. (2010) What do people look at when they watch stereoscopic movies? *IS&T/SPIE’s International Symposium on Electronic Imaging: Science and Technology. Stereoscopic Displays and Applications XXI*, 18.-21.1.2010, San Jose, California, USA. Proceedings of SPIE, Vol. 7524.
- Mital, P.K., Smith, T. J., Hill, R. and Henderson, J. M. (2011) Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1), 5-24
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634-1640.

Münsterberg, H. (1970). *The film: a psychological study; The silent photoplay in 1916*. New York, USA: D. Appleton and Company.

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

Nuthmann, A., Smith, T.J., Engbert, R. and Henderson, J.M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, 117(2), 382-405.

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

Smith, T. J. (in press) Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring Cognition at the Movies*. New York: Oxford University Press.

Smith, T. J. (2012) The Attentional Theory of Cinematic Continuity, *Projections: The Journal for Movies and the Mind*. 6(1), 1-27.

Smith, T. J., Levin, D. T. & Cutting, J. (2012) A Window on Reality: Perceiving Edited Moving Images. *Current Directions in Psychological Science*. 21: 101-106