Extending AToCC: a reply

Tim J. Smith

31/12/11

Word count = 3,019 words

The intention of my paper, *The Attentional Theory of Cinematic Continuity* (AToCC; this volume) was to bring together disparate lines of research on cinematic continuity and make an intentionally provocative first stab at theorising a cognitive foundation for the continuity editing conventions that pervade film. If our understanding of film cognition is to progress, clear hypotheses need to be proposed that can be empirically tested and theoretically challenged. To this end, the high quality and variety of the six responses that constitute the rest of this volume are hugely encouraging. My only regret is that I will not be able to deal with each of their insightful comments in the space permitted.

Seeing through film

Several of the respondents (Turvey, Freeland, and Smith) question the confidence with which I borrow from the "ecological" view of film perception (e.g. Anderson, 1996) as they claim that the principles of continuity editing fail

mirror the way we attend to and perceive real-world scenes. My argument is that *active vision* deployed in service of an inquisitive viewer looking to seek out information of interest, but restricted by a limited capacity attentional, visual and memory system, invalidates many of the incompatibilities previously identified. However, Turvey rightly points out that, while I state in my original article that a cognitive theory of continuity must specify how we overcome these incompatibilities, I fail to do so, only indirectly dealing with the incompatibilities as I explain the cognitive foundations of several key editing rules. Let me address this concern by replying to the problem as quoted by Turvey: "*in real life every experience or chain of experiences is enacted for every observer in an uninterrupted spatial and temporal sequence*" whereas this is "*not so in film*" due to the fact that "*the time that is being photographed may be interrupted at any point*" by editing (Arnheim, 1957; pp. 20-21).

Perception, not sensation

Before addressing this problem, I should emphasize that AToCC proposes that continuity editing mirrors real-world *perception* and not *sensation*; the incompatibility problem as stated by Arnheim is phrased in terms of audiovisual sensation and not our perception of it. I agree that the physical form of film bears very little resemblance to the light projected on to a static[i] retina by a spatiotemporally continuous real-world scene. However, we do not have direct access to this sensory stimulation. Our experience of an audiovisual environment is mediated by perception. For film to mirror our experience of an audiovisual

scene it must, therefore approximate our perception of that scene rather than the direct sensation.

However, reframing the problem in terms of perception does not negate the argument that we normally perceive real-world scenes as spatiotemporally continuous and editing violates this assumption by involuntarily shifting our viewpoint in space and time. How does AToCC explain this apparent incompatibility?

Does active vision require agency?

In her response to my article, Cynthia Freeland emphasizes the issue of *agency* and wonders how continuity editing can mirror natural perception without the viewer voluntarily influencing the changes in viewpoint. I argue that the instantaneous change in viewpoint presented across a cut does not present a problem for our visual system because such changes occur 3-5 times every second when we perform a saccadic eye movement. Although a change in shot may also distort perspective on a scene or location of the observer/camera, these distortions pass unnoticed (Hollingworth & Henderson, 2004) as long as what we are interested in is instantly recognisable after the cut. I state that continuity "*is about enabling the viewer to shift their attention to the audiovisual details currently relevant to them and the narrative*" (pp. 15). A cut that preserves *a priori continuity* must cue attention prior to a cut so that the change in viewpoint mirrors our desired attentional shift. This makes the viewer a collaborator in the change in viewpoint, giving the illusion of agency.

However, Freeland's concern about the absence of agency extends beyond the motivation for a change in viewpoint to the issue of which neural pathways process the changing visual input.. She borrows from Matthen (2005) the distinction between "descriptive" vision and "motion-guiding" vision or more commonly referred to "vision for action" (Milner & Goodale, 1995). The former is used to represent visual scenes from a distance, without the intention of interaction whereas the latter represents objects and the scene relative to the viewer in preparation for the motor plans required for interaction. These two types of vision utilize distinct but interacting visual pathways in the brain: the ventral ("What") and dorsal ("Where") pathways, respectively. Given that viewers can never inflict their will on a film, Freeland argues that we must adopt anunnatural purely descriptive viewing mode that only utilizes the ventral pathway. Freeland's argument implies that she believes the absence of vision for action distances our visual experience of film from reality to an extent that it invalidates my argument that continuity editing mirrors the way we perceive reality. I would argue against this interpretation as it neglects the interaction between the two visual pathways and the fact that vision for action is normally preceded by descriptive vision. Motor actions are often preceded by "look ahead" fixations that identify relevant objects, their location, and their predicted movements relative to the agent (Land & Taler, 2009). For example, when driving a car we fixate the section of the road 1 second ahead of our current location irrespective of driving speed (Land & Lee, 1994). These anticipatory glances are thought to enable us to build a representation of the space within

which we will act, such as the curvature of the bend. The motor act of steering the wheel occurs after our gaze has sought out the relevant information.

I would argue that continuity editing mirrors the first two stages of attention and motor planning but replaces the final stage of motor action with a cut. When observing a real-world scene such as a conversation we may find our viewpoint of a relevant part of the scene obscured either by being out of our field of view, necessitating a head rotation or occluded by other objects, necessitating a change in viewing position (see Turvey's response in relation to matched-exit/entrances). The motor plans to change viewpoint will be preceded by anticipatory shifts of gaze and/or attention in the direction of the desired object. When such a scene is represented via continuity editing, our anticipatory shifts of attention remain and we may also prepare the motor plans necessary to execute the change in viewpoint (e.g. rotating our heads to look off screen) but the cut beats us to it by providing the desired shift in viewpoint. We have engaged in two thirds of the cognitive processes involved in the volitional change in viewpoint which may be enough to trick us into believing we were responsible for the shift or, at least to not invalidate our *a priori* assumptions of spatiotemporal continuity. As Freeland concludes, the extent to which the viewer is active across a cut and the two visual pathways are engaged should be a topic of future research. The most promising techniques for answering this question will be neuroimaging (e.g. functional Magnetic Resonance Imaging; fMRI) due to its ability to directly identify the activity of the two neural pathways during film viewing.

Adapting to film

Turvey and Freeland seem to believe that the absence of the final stage of actively attending to an object in the real-world, i.e. the motor act of saccading or rotating our heads to locate the target, negates the ecological view of film perception. I do not believe this is the case. . There is no default mode in which we perceive the real-world. Our visual system is constantly adapting to the demands of each environment and our intentions within it. This adaptation can be physiological, e.g. pupil dilation, binocular vergence and accommodation of the lens to suit the luminance and depth of an image; motor, e.g. modifying the amplitude of saccades and their frequency to the density of information; or perceptual, e.g. changing how we map sensory input to representation. Later in his response, Turvey agrees that trying to compare the way we perceive the real-world to that of film is a fool's errand as both are highly variable and open to a high degree of interpretation.

The most striking example of perceptual adaptation can be demonstrated using prism glasses. Hermann von Helmholtz first demonstrated this phenomenon at the end of the 19th century when he showed that wearing prism glasses inverting the visual world initially made interaction with the world impossible but the wearers quickly adapted. This adaptation was subsequently shown to be due to neural plasticity: the changing of neural pathways and enlisting of existing pathways for different ends. Such plasticity is critical for our survival in ever changing environments. Perceptual adaptation and rapid learning of the rules governing a particular sensory experience are natural

aspects of real-world cognition. Without them infants would never learn how to stand, walk or talk. Adapting to the constraints of a new audiovisual environment such as the internal logic of an Avant Garde movie is not, as Freeland states in her response, "unnatural" but a critical component of real-world cognition.

I believe that the way in which continuity editing approximates our normally volitional changes in viewpoint within a scene may constitute a form of adaptation. Even without much prior exposure to film we rapidly understand the physical constraints of the screen and selectively inhibit the final stage of actively orienting to objects within the depicted scene, handing control of our head rotations and viewpoint changes over to the filmmaker instead. Such immediate adaptation may partly explain the anthropological evidence that viewers who have never previously been exposed to edited moving-images are still able to comprehend the events depicted (cited in Messaris' response). Schwan and Ildirar (2010) have recently presented similar evidence of immediate adaptation to film by naïve film viewers in a remote region of Turkey. Schwan and Ildirar constructed short edited sequences using a range of continuity rules such as establishing shots, shot/reverse-shot, crosscutting, and temporal ellipsis. They found that naïve viewers were able to comprehend edited sequences as long as they depicted a familiar and clear line of action. Conventions that did not contain continuity of action, such as an establishing shot of a house followed by a man inside the house required extensive prior exposure and would be described by naïve viewers as seperate scenes. This evidence supports my proposal in AToCC that sensory inauthencity such as that

described by Arnheim is irrelevant as long as viewers can attend to what is of interest across the cut and perceive the relevant features of a scene. Focused perceptual authenticity trumps sensory inauthenticity.

The curious case of the missing comic continuity[ii]

Adaptation of existing perceptual skills to the physical constraints of the medium also explains Greg Smith's comment about the differences between the conventions used to present spatiotemporally continuous scenes in film and comics. Comics do not always adhere to the 180 degree rule because viewer gaze is constrained by reading direction (left to right, top to bottom in Western comics) and the need to fit content to the page. Comics borrow heavily from cinema, often using establishing shots, shot/reverse-shot sequences and the analytical breakdown of a scene via frames (McCloud, 1994).However, Greg Smith is correct to identify that cinematic conventions that preserve a line of action (e.g. 180 degree rule) do not fit so comfortably with the physical constraints of reading direction and are often sacrificed. Rather than posing a problem for the ecological component of AToCC I believe these medium specific differences confirm my emphasis on the "continuity of attention": being able to shift attention to the features of a scene that are of interest to the viewer/reader. Comics accomplish this by isolating spatiotemporal snapshots of key moments of a scene and implying the spatiotemporal relationships between frames. This reduces a scene down to its communicative essence while accommodating the physical constraints of the medium. A novel uses a similar but more extreme form of reduction via a symbolic form of representation and relies more heavily

on the viewer to represent the communicated events. We can view literature and film as two extremes on a spectrum of immediacy with comics somewhere in between. The intention of each mode of storytelling may be similar: communicate the essential events of a story within the constraints of the medium. However, the sensory immediacy of film means that this must be accomplished by accommodating our natural perceptual tendencies such as orienting to motion, out of sight sounds or following gaze cues.

The primacy of narrative action

Returning to the rephrased incompatibility between film and reality posed at the beginning of this article, how does AToCC explain spatiotemporal ellipsis? Several of the respondents (Turvey, Freeland, Smith, Rogers, Levin and Hymel) praised AToCC for the depth with which it tackled issues of how we perceive continuity across cuts that have the intention of preserving spatiotemporal continuity, e.g. match-Action cuts. However, they questioned how it might be extended to deal with "the richer more meaningful experiences" (Rogers, pp. 2) that are the typical realm of film theory such as emotion, narrative, and aesthetics. Sheena Rogers' three levels of aesthetic experience -- sense and soma, art and affect, truth and transcendence -- may capture the full range of cinematic experiences but I believe they are beyond the scope of a single cognitive theory of film like AToCC. However, I believe narrative comprehension (within Roger's second level) is a tenable extension of AToCC. As correctly hypothesized by Freeland "Following the flow of visual continuity, a viewer is probably at the same time generating hypotheses about both the smaller and the

larger stories within which the visual episodes fit" (pp. 8). I briefly discuss such higher-order hypotheses in the section *Complex cuts and combination of cues* (pp. 28) but a full expansion of how I believe AToCC can operate at the level of narrative would require considerably more space than is permitted.

In brief, I believe that the three stages of AToCC presented in relation to spatiotemporal continuity – cuing attention pre-cut, formulating perceptual inquiries, and matching minimal expectations post-cut – also operate across cuts with action and narrative continuity. For example, establishing a rhythm of cross-cutting between two parallel actions can allow viewers to prepare for the shift in attention to the new action and formulate expectations about what will be presented after the cut. A chase sequence spanning many locations and points in time focuses viewer attention on the action and does not require precise spatiotemporal continuity. The sequence of alternating matched-exit/entrance cuts need only to present snapshots of the action in which critical events happen such as the protagonist doubling back or the pursuer catching up. Viewer attention is focused on the plight of the protagonist and they only need to represent sufficient information to comprehend this event such as proximity between protagonist and pursuer.

Evidence for the primacy of narrative action comes from the visual anthropology studies cited earlier (e.g. Schwan & Ildirar, 2010) and the anomalous edit blindness for within-scene cuts (Smith & Henderson, 2008). In my edit blindness study, alongside the finding that a third of all match-action cuts were missed during an explicit cut detection task, I also found that a quarter

of all within-scene cuts were missed. These cuts were not preceded by an onset of motion or an actor's gaze shift. However, within-scene cuts resulted in significantly greater edit blindness than between-scene cuts or gaze-cue cuts. What is continuous across most within-scene cuts is action, whether it be dialogue, actor movement or interaction. fMRI evidence from Magliano and Zacks (2011) suggests that cuts with action continuity may be processed by different brain regions than cuts with spatiotemporal continuity and spatial updating may be suppressed in favor of attending to the action. Whether cuts with action continuity utilize the three stages of AToCC and what the consequence of violating each stage is on the perception of action continuity must be the topic of future empirical investigations.

The fallacy of prediction

Finally, I commend Dan Levin and Alicia Hymel's response for the depth to which they engaged with the implications of AToCC. Levin and Hymel criticize AToCC for its over-emphasis on prediction. They rightly question whether "it is really necessary to be constantly predicting the near future?...as the activity is likely to be computationally and representationally intensive." AToCC borrowed this emphasis on prediction from Jeff Zack's event segmentation theory (EST) which states that we perceive boundaries between events in natural human activities when our ability to predict the future form of the activity drops below a certain threshold (Magliano & Zacks, 2011). AToCC takes this idea and incorporates it into the second and final stage of continuity perception:

formulating expectations and checking expectations post-cut. However, Levin and Hymel's evidence that participants failed to notice the reordering of actions in an edited sequence suggests that prediction doesn't happen by default and that *postdiction*, i.e. prediction after the fact, might actually be more common. In AToCC, postdiction would allow the perception of *a posteriori* continuity but only after the perception of *a priori* discontinuity. If Levin and Hymel are correct, this may mean that *a posteriori* continuity is more common than *a priori* continuity even within films strictly adhering to the Hollywood style. This hypothesis should be tested in two ways 1) statistical analysis of the frequency of cuts in existing movies that satisfy the conditions for *a priori* continuity outlined in AToCC, and 2) further empirical investigations to identify the perceptual consequences of cuts that do or do not meet these conditions.

In conclusion, I am very grateful to the respondents for drawing attention to assumptions of AToCC and necessary extensions that I omitted to address in the initial article. I hope I have provided insightful replies to these comments and suggested ways in which AToCC may be extended in the future. Film is an ever changing, constantly amazing, and beautiful medium that goes right to the core of what makes us human. I hope that the theory I have presented here and in the main article expands our understanding of the cognitive processes involved in film perception without losing any of this majesty.

References

Arnheim, R. 1957. *Film as Art*. Berkeley, California: University of California Press.

Hollingworth, A., & Henderson, J. M. 2004. "Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory." *Perception & Psychophysics 66*: 800-807.

Land, M. F. & Lee, D. N. 1994. "Where We Look When We Steer." *Nature* 369: 742-744

Magliano, J.P. & Zacks, J.M. 2011. "The impact of continuity editing in narrative film on event segmentation." *Cognitive Science 35*: 1-29.

Matthen, M. 2005. *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Oxford University Press.

McCloud, S. *Understanding Comics: The Invisible Art*. Harper Collins, 1994.

Milner, A.D., & Goodale, M.A. 1995. *The Visual Brain in Action.* Oxford: Oxford University Press.

Schwan, S., & Ildirar, S. 2010. "Watching film for the first time: How adult viewers interpret perceptual disconuities in film." *Psychological Science* 21(7): 1-7.

Smith, T. J., & Henderson, J. M. 2008. “Edit Blindness: The Relationship Between Attention and Global Change Blindness in Dynamic Scenes.” *Journal of Eye Movement Research* 2(2): 1-17.

Endnotes

[i] I specify a "static" retina as I will argue later that the addition of eye movements, shifting the image on the retina and changing perspective on the scene is the first way in which our perception of the real-world begins to approximate that of edited moving-images.

[ii] By “continuity”, I am referring to the same spatiotemporal continuity discussed in relation to film. The term is confusing in the context of comics as “continuity” is more commonly used to refer to consistency in the story universe belonging to each of the main comic book publishers (e.g. DC and Marvel). This form of continuity ensures that when Batman kills a nemesis in one issue the same nemesis doesn’t crop up alive in a later issue of Spiderman (unless it is in an “alternative universe” plot arc!). Such universe continuity has been absent from comic book movie adaptations to date but this will be addressed in *The Avengers* (2011).