

Attention, Blending, and Suspense in Classic and Experimental Film

Todd Oakley and Vera Tobin
Case Western Reserve University

In *Blending and the Study of Narrative* (2012). Eds. Ralf Schneider and Marcus Hartner. Berlin: Mouton de Gruyter. 57-83.

1. Introduction: Suspense and Film Art

It sometimes seems that every academic discussion of suspense in narrative, and especially of suspense in film, must reference Alfred Hitchcock's account of the phenomenon (Truffaut 1967 [1983]), and this chapter will be no exception. Hitchcock claimed that suspense arises when viewers have—or believe that they have—privileged information about a forthcoming undesirable event and are powerless to intervene. He presents an illustrative example:

We are now having a very innocent little chat. Let us suppose that there is a bomb underneath this table between us. Nothing happens, and then all of a sudden, "Boom!" There is an explosion. The public is surprised, but prior to this surprise, it has been an absolutely ordinary scene, of no special consequence. Now, let us take a *suspense* situation. The bomb is underneath the table and the public *knows* it, probably because they have seen the anarchist place it there. The public is *aware* that the bomb is going to explode at one o'clock and that there is a clock in the décor. The public can see that it is a quarter to one. In these conditions this same innocuous conversation becomes fascination because the public is participating in the scene. The audience is longing to warn the characters on the screen: "You shouldn't be talking about such trivial matters. There's a bomb beneath you and it's about to explode!" (p. 73)

In Hitchcock's formulation, suspense is a function of information management and audience anticipation. The tension arises from a discrepancy between what viewers know (or believe that they know) about what is going to happen and characters' lack of awareness regarding these circumstances. It seems to be a particular species of dramatic irony, constrained by circumstances involving some kind of looming danger and uncertainty about whether the characters will respond in time to avert it.

Cognitive approaches often focus on how this definition hinges on the question of what viewers *know* and what they merely *anticipate* or expect. The psychological, philosophical, and cognitive literature on suspense has thus been especially concerned with what is sometimes called the "paradox of suspense" or the "anomalous experience of suspense" (e.g. Gerrig 1989, Brewer 1996, Yanal 1996, Carroll 2001). The question that plagues these studies is why people should report feeling suspense when they are seeing a film for a second time, or when genre conventions

make it trivial to predict the outcome of some sequence. In other words, if suspense involves uncertainty, why should people feel suspense when they know what's going to happen?

In the present study, we are more interested in another aspect of the scene that Hitchcock describes: namely, the extent to which its suspense—and the information management required to generate that suspense—seems to be the product specifically of social cognition. That is, it is not merely a matter of tracking information about a sequence of events, but of processing and applying social information, managing multiple perspectives, and ascribing mental states to real and imagined participants in the narrative scene. Film is vigorously, inescapably intersubjective. It engages our everyday systems for making sense of other people's behaviors in intentional terms and for thinking interactively.

At the same time, we would argue that suspense is more than the product of "mentalizing," or simply knowing what a given character does and does not know. This epistemic stance only scratches the surface of what is happening cognitively. Hitchcock's canonical scene of suspense is indeed a central example of the experience, but it has little to say about the kind of suspense generated in a film like Michael Snow's 1967 experimental feature *Wavelength*, which consists of a forty-five minute zoom shot in a single room, during which character actions occur only briefly and peripherally. But we will claim that this film does generate suspense, and can explain what this construction has in common with the effects associated with traditional suspense narratives like Hitchcock's. In addition, any account of suspense must deal with failed attempts to generate it, a predicament common with many so-called "B" movies.

Noël Carroll (2003, p. 26) observes that "...filmic communication, to a large extent, depends upon the filmmaker's control of the audience's attention." We agree wholeheartedly. The question, then, is why moving images control audience attention the way that they do and how that control affects the audience. To explain how suspense emerges from the interaction of narrative conventions and our natural cognitive responses, we will need to look at the contributions of two important cognitive systems that underlie the construction of meaning: (1) joint attention, in which two or more people are both focused on an external object and mutually aware of this shared focus; and (2) conceptual integration, in which elements and vital relations are selectively projected from multiple inputs, and processes of composition, completion, and elaboration give rise to new emergent structure in the blend. Each without the other fails to provide an adequate cognitive model of the phenomenon of suspense as it plays out in real time.

Suspense is a product of manipulations of blended joint attentional scenarios, or what we will call "joint attentional blends". The ingredients of suspense are the ingredients of human cognition generally. Film is a medium that is especially amenable to manipulations of our investment in the joint attentional triangle, because it so tightly integrates the gaze of spectator and camera. By building on this

fundamental integration, filmic conventions generate suspense by constructing triangles of joint attention in which the attentions of the filmmaker, or the camera-eye, overtly overpower the agency of the viewer. They engage our natural systems for social cognition while frustrating our desire to complete the joint attentional triangle.

We use this model for the analysis of two classic Hollywood films, Hitchcock's *Notorious* (1946) and Wells's *Touch of Evil* (1958), and one experimental film, Michael Snow's *Wavelength* (1967), and finally discuss what happens when films fail to engage these systems effectively, so that their attempts at generating suspense fall flat.

2. Joint Attention and Conceptual Integration

The ability to engage in sustained scenes of joint attention is a crucial ingredient for the development of language and communication and is part of a larger suite of social cognitive skills, including imitation and intention reading.

Some time around the end of their first year, children begin to engage in a number of new activities that involve not just themselves and an object nor just themselves and another person, but themselves, an adult, and objects in their environment toward which both infant and adult direct their attention. Before then, neurotypical infants can engage in dyadic behaviors—interactions between themselves and an adult, or themselves and an object—but as they approach their first birthday, they begin flexibly and reliably to coordinate their activities within a referential triangle of themselves, an adult, and an object or event towards which they're sharing attention.

In its earliest and most fundamental forms, this interaction involves reference to an object that is directly mutually perceptible to the participants in the communicative act. There are three crucial elements in a joint attentional scene: (1) the ego, (2) the other, and (3) some third object to which they coordinate their attention. Joint attention also involves a shared intentional relation to the world, and the joint attentional frame is defined through participants' shared understanding of goal-directed activities in which they are jointly engaged. In full-fledged collaborative engagement, participants' joint attentional frames subtend interactions directed toward shared goals, with coordinated action plans.

The emergence of coordinated joint activities and their concurrent joint attentional scenes around the end of the first year of life are well documented, and early signs of linguistic communication co-occur with the emergence of these joint attentional activities (Hay 1979; Bruner 1983; Bakeman and Adamson 1984). Some findings indicate that the amount of time spent in joint attention activities in the first year is predictive of vocabulary at 18 months (Tomasello and Todd 1983).

Humans are experts at understanding the behavior of themselves and others in terms of beliefs, desires, and intentions. We can ascribe mental states (which we recognize may differ from our own) to other people; we can predict how these mental states might change in response to various things; and we can use these ascribed mental states to predict or make sense of behavior. All of this comprises our everyday “mind reading” (Baron-Cohen 1995), sometimes called “theory of mind” (Premack and Woodruff 1978) or “concept of mind”. But because we don’t have the advantage of direct telepathic access to one another’s thoughts, we have to make do with inferences based on a combination of our background knowledge and our observations of what people say and do, and to do this we are especially attuned to markers of attention in general, and visual attention in particular.

You can deliberately draw another person’s attention to an object by remarking on it, pointing to it, waving it in her face, holding it up to her line of sight, offering it to her, or taking her to a new place so she can observe the object there, and she can reliably understand these actions as intentional behaviors aimed at directing her attention (see Oakley 2009 for a typology of attention directing linguistic devices). Humans are the only animals to engage in these kinds of social activities habitually in their natural habitat (see Tomasello 1999 for an overview of relevant studies). Language and the visual and filmic arts all provide an array of sophisticated tools for directing and coordinating attention. Joint attentional scenes comprise the crucial ingredient of the cultural ecology that subtends language and the visual and filmic arts. Without it, these symbolic activities disappear.

As well as responding to events where others explicitly bring things to our attention, we also keep an eye on what others are attending to as they go along. Following the gaze of those around us is an adaptive behavior. Doing it reflexively is a good way of increasing the odds of locating important things like food and predators. Mature human social cognition, in which we understand other people as having mental states like ourselves, invests these movements with special significance. We understand that other people turn their heads, for example, in the course of intentionally shifting their gaze towards external objects that interest them. Armed with this understanding, when you watch another person look at something, you aren’t simply making it more likely that you’ll see something of general interest to you. You’ve gained information about the degree to which she is aware of it.

Films, especially narrative films, present a number of overlapping joint attentional scenarios that are the product of conceptual integration. The gaze of the spectator is tightly integrated with the viewpoint of the camera, which does not merely direct the attentions of the former but in fact controls them to a significant degree (as long as the spectator is watching the film at all; once she is not, she has largely relinquished the role of spectator, until she re-aligns her gaze with the camera’s). This integration presents a number of opportunities to force the spectator into nearly but not-quite completed attentional triangles in the blend. There may be characters who refuse to meet the camera’s eye or to follow its gaze to some

important object; or the camera itself may refuse to focus on the object to which the spectator would like to direct her own attention.

Meanwhile, the objects of the camera's "attentions" are not genuinely present to the spectator. Their reality and immediacy obtain in the blend, rather than in the spectator's own space. Whenever the joint attentional triangle takes the spectator as one of its participants, its jointness is always fictive at best. The apparent attentional intelligence behind the camera's movements is itself the product of a blend. The contributions of many different participants direct the camera, its motions, what is seen and what is not. In the blend, however, these individuals are blended into a single agent.

The moving images of film and the suite of conventions by which those images are organized generate especially vivid joint attentional compressions. The vividness of these blends is crucial to the impulse to shout at the screen that Hitchcock describes. As Fauconnier and Turner often remind us (e.g. 2002, p. 260), we aren't deluded, but the blend is powerfully compelling. Whether that compelling blend gives rise to the sensation of suspense, and the active "longing to warn the characters on the screen," depends largely upon when the closing of the triangle in these blends is or is not imminent.

3. Camera Movement and Attentional Triangles in *Touch of Evil*

First appearing in popular films around 1914, the tracking (or dolly) shot uses a camera mounted on a track, along which it can "dolly in" or "dolly out" on stationary subjects for emphasis, or "dolly with" a moving subject. A related long shot technique is known as the crane shot, which allows the camera to move on along both a vertical and horizontal axes, to produce complex "continuity movements" from aerial vantage points.

Our first filmic example is the three minute and thirty second opening crane shot of Orson Wells's film noir masterpiece *Touch of Evil* (1958), widely regarded as one of greatest long takes in Hollywood Era film. The sequence takes its viewers from the planting of a bomb and its timing device in the trunk of a car, tracking the booby-trapped sedan as it wends its way through the Mexican border town with pedestrians, traffic police, street vendors, and boarder guards going about their business, across the border to its final resting place. The car is, in our terms, a "vessel of suspense." Viewers share knowledge that the car is going to explode in three minutes but they do not know where or when it will happen and what the collateral damage will be. Who in addition to the occupants of the booby-trapped car will be killed: a policeman, a border guard, or street vendor, or the couple walking toward the US side of the boarder? Wells' cinematic vision ratchets up the tension at several places in this sequence and there is a strong conventional relationship between camera movement and degree of suspense that can be accounted for by joint-attention within a network of related mental spaces.

The camera's gaze is an indicator of significance, for, in our model, the camera functions as a proxy for the conversational partner or interlocutor in a joint-attentional scene. More specifically, it operates much like an adult-caretaker in scenes of child-adult interaction, as it is usually the adult caretaker who initiates conversation and joint activities (at least initially). But with cinema, even more than perhaps any other narrative format, viewers only see what the directing viewpoint—the camera—lets them, and thus shot sequences and camera movements have a determining effect on what we regard as significant and how we conceptually integrate relevant information into compelling, and in this case, suspenseful, scenes.

Through its movements, the camera can index and bracket the viewer's attention, and at times, as with extended tracking shots, the viewer can become highly conscious of the fact that his attention is being directed and manipulated. The viewer, however acutely aware of this manipulation, has few options but to yield to the camera's will, as is palpably the case with this opening scene. The camera follows the booby-trapped car through town (shot T1) then it seems to shift focus to the couple walking toward the border (T2), whom we learn are newly married, in effect bracketing the car from view, until the strolling couple and lurching car are side by side in the same frame (T3). At this moment, the car and the couple are the clear objects of the camera's gaze, and the ensuing action and dialogue provides viewers with additional information about the couple, namely that the husband is a Mexican lawyer who has prosecuted a drug kingpin of the feared Grande family and his wife is a US citizen born in Philadelphia. The tracking shot also allows the vessel of suspense to move off camera while maintaining the focus on Miguel and Jenny Vargas as they kiss. The explosion happens on the US side of the border, an event that sets in motion the events of the movie.



Figure 1: Shot T1



Figure 2: Shot T2



Figure 3: Shot T3

The mental format of this opening sequence can be understood in terms of a conceptual integration network in which both the content and the terms of its presentation are modeled simultaneously and unfolding online and in real time. The triangulation of attention to a salient third object on the strict terms of the camera's gaze is what makes cinematic suspense both a commonplace phenomenon, building on a standard mental format of joint-attention and intentional reading, and an exceptional narratological technology that carefully polices the means, manner, and condition of the viewer's attention, so that the viewer lacks the ordinary commensurate ability to engage in imitative reversal, i.e., the opportunity to do what the "conversational partner" does and genuinely coordinating attention with a co-participant.

The mental spaces format of narrative suspense is outlined below. As a preliminary measure, we make explicit our definition of a mental spaces, a conceptual integration network, a conceptual blend, as well as say a few words about the cognitive status of the network being presented in these pages.

First, mental spaces consist of scenes or scenarios or facets of scenes and scenarios that are active in working memory and which are structured by frames and semantic domains and which often develop vantage points, perspectives, and temporal dimensions. Second, a network of mental spaces typically follows a mental format in which a Presentation and Reference Space contribute structure to a third Virtual or Integrated space. This network of mental spaces in turn issues from about feeds back into what we call the Ground, following Clark 1996. This Ground reflects the ontological status of the relevant actors in a film-viewing scenario. In this case, it follows the asynchronous and asymmetrical format of the filmmaker and film spectator. The network as configured here represents what we like to call "center stage" cognition, this in contrast to "back stage cognition" posited by Fauconnier and Turner (2002). By center stage cognition, we mean the network of scenes and scenarios accessible to immediate consciousness—namely the events as they are unfolding onscreen, the formal conditions of their disclosure to viewers, who are often but not continuously conscious of them, and the emergent properties that result from integrating these two dimensions with the typical product of suspense and anticipation.

The Presentation space corresponds to the camera gaze and the “caretaker” role in the joint attention scene. In this space, the viewer triangulates attention to the restricted scope of and terms of presentation offered by the camera gaze. The camera literally “behaves” and the viewer responds to that behavior as intentionally communicative. In the case of Wells’ film, this attention disposition is “field-based” and external to the events depicted. That is to say that the viewer and camera observe the events unfolding from an observer rather than character point of view. Simply put, this means that the camera gaze maintains a vantage point such that everything beyond its scope counts as “off stage” with respect to the represented world of action. We argue that tracking shots not attributed to any focalized character assiduously maintain this observer point of view. (As we will argue, the continuity editing of other films can interpolate the viewer into the on stage business.) One can legitimately ask at this point, what is the difference between a Presentation space and Ground? Isn’t the Ground the space of negotiation between viewer and filmmaker that you are positing as the very basis for this joint attention enterprise?

The answer is that the Ground simply specifies a generalized ontological category of cinematic engagement that is part of the taken-for-granted features of symbolic interaction that are important features of theorizing the reception of any extended discursive and symbolic engagement. The Presentation space, on the other hand, models what we might call (after Edelman 1989) the “remembered present”—the moment-by-moment interaction between the camera’s gaze and the viewer with respect to the conventional means of presenting the dramatic action on the screen. Just as an interlocutor can be conscious of, critical of, or otherwise respond to the communicative status of a speaker’s utterance (e.g., what do you mean by “you people?”), the viewer can be conscious of, engage with, and be critical of the terms by which the camera discloses information. For instance, one minute into the tracking shot, one of the authors became acutely aware of the limited scope of the tracking shot and “wanted” the camera to alter its gaze, just for context and variety. The upshot is that the Presentation space in our network covers the mental scenarios associated with the production of the story. Other versions of this space include the actors’ identities, such as “Vivian Leigh is kissing Charleton Heston,” where the perspective of the actors’ action is understood from the perspective of the conditions of enactment and not the story world.

The Reference space corresponds to the staged business or story world being depicted on screen. Viewers understand this as an opening scene for the ensuing drama, and thus everything appearing therein has significance even if the viewer cannot infer the specific significance of any one thing occurring therein. The viewer learns immediately that a bomb was planted in the trunk of a car, and that a couple got in the car and if they do not get out in three minutes, they will be dead.

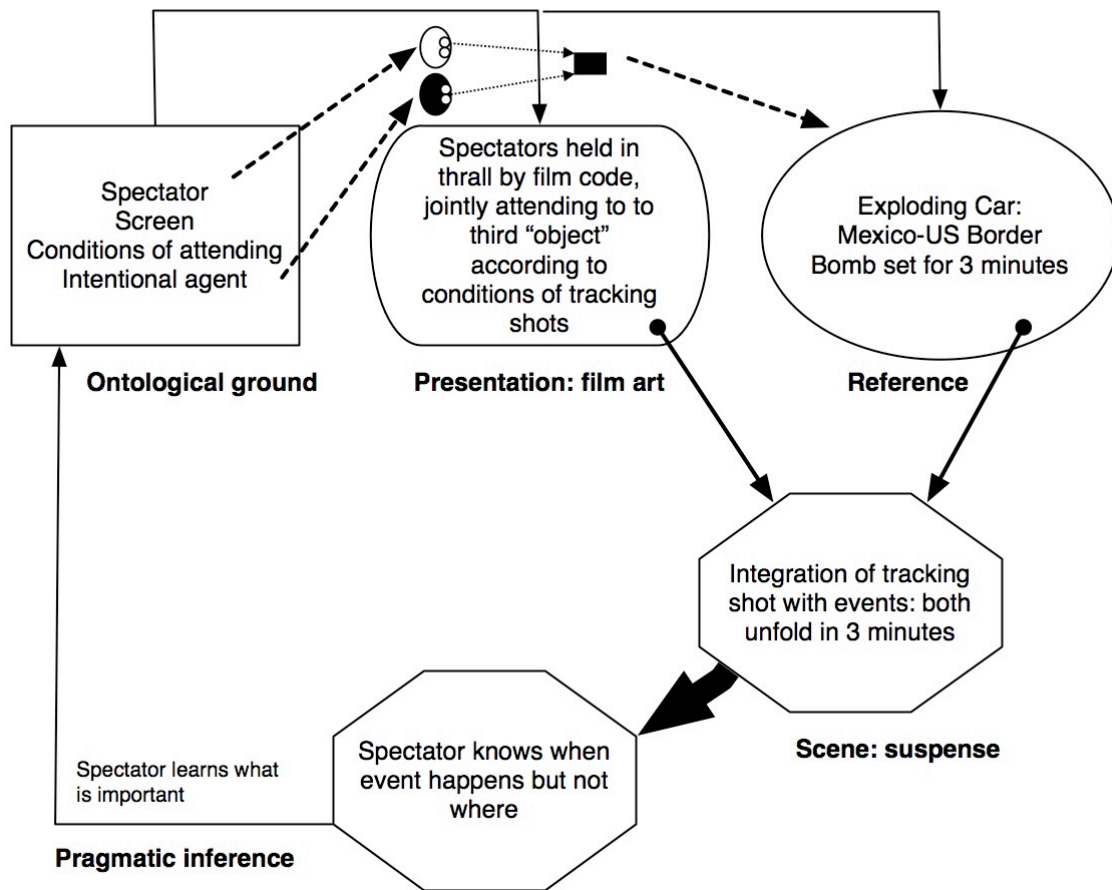


Figure 4: Blending diagram for the opening sequence in *Touch of Evil*

The viewer knows that an event of human scale significance is going to take place in three minutes. What they do not know is where it will happen. The camera, however, will follow objects of interest. When the camera tracks the strolling couple, the viewer confers upon them significance simply by virtue of the camera gaze. The audience then feels that the couple walking on the street is in as much peril as the couple in the car. This inference is the result of the conceptual integration of the elements from the Presentation and Reference spaces, hence what we call the Virtual space of suspense. Thus, at 2:20 to 3:18, when the newlyweds and the couple in the car occupy are both stopped at the border, the sense of suspense is at its height. It is at this moment especially where the tension is fever pitch, where the focus of attention is sustained on the principal character and the rigged automobile, when we know the couple in the car are going to die and, perhaps, feel slightly relieved when they drive off screen and the camera focuses in again on the Vargases.

4. The Continuity System and Attentional Triangles in *Notorious*

Our second example returns to Hitchcock with a scene from his 1946 film *Notorious*, in which Ingrid Bergman plays Alicia Huberman, the American daughter of a convicted Nazi spy. She is brittle, proud, in desperate need of love and acceptance, and has been recruited to infiltrate a group of her father's former confederates who have reconvened in Brazil. Cary Grant is her handler, an American agent called T. R. Devlin. Naturally they fall in love, but Devlin is held back by his knowledge of her licentious past, and Alicia by her pride.

Their relationship is further complicated when Devlin's superiors decide that Alicia must seduce one of the confederates, Alex Sebastian (Claude Rains). Sebastian is patently head over heels in love with Alicia, and makes his feelings clear in a way that the rigid Devlin cannot bring himself to do. Eventually Alicia, convinced that Devlin has never loved her, even agrees to marry Sebastian, the better to gain his trust and close access to his secret machinations.

When Sebastian eventually discovers Alicia's treachery, he must stop her from exposing the plot (revolving around an unusually prescient MacGuffin of pilfered uranium) but also avoid exposing his own misstep to his Nazi associates, who will not be tolerant of such a lapse in judgment. Fortunately he is in possession of a ruthless and controlling mother, who decides that Alicia "could become ill. And remain ill for a time, until..."

There follow a number of tightly wrought scenes over the course of which the Sebastians bring Alicia to the brink of death via slow poisoning. In one of these scenes, the movements of a coffee cup serve as the focal point of two and a half minutes of mounting tension, culminating at last in Alicia's sudden awareness of the cup's significance, just in time for her to succumb to the poison and collapse. Throughout this scene, the camera returns obsessively to the cup, tracking it from Madame Sebastian to Alicia, keeping it in the foreground of each shot on Alicia, watching it travel from the table to her lips and back. Meanwhile, everyone else is infuriatingly unwilling or unable to even glance at it.

The tight focus on the coffee cup both establishes its significance—we never see the poison or any discussion of how it will be administered—and makes it the focal point of a deliciously prolonged sequence of building tension. The spine of the suspense would seem to be entirely situational: Alicia and her poisoned coffee are together at the table. An undesirable event, the moment when she drinks it, is imminent.

The suspense, however, arises not only from this anticipation of events, but from the creation and organization of joint attentional triangles surrounding a crucial object, via the strategic deployment of several highly conventionalized editing techniques. The scene presents several potential attentional triangles where the putative participants can be characters as well as the camera eye. Even as the cup is being

made deliberately salient for us in the triangle of filmmaker-viewer-object, several instances of intentional inattention on the part of participating characters intervene. The scene's dramatic tension is driven by the way that characters resist attending to the crucial item, thwarting the consummation of the attentional triangle and generating suspenseful narrative tension. This sequence is managed via a number of editing techniques that conventionally facilitate spatial, temporal, and psychological continuity across shots.

- Until the scene's climactic moment, Alicia barely looks at the cup. She pays no visible attention when it is placed before—prominently placed in the extreme foreground of the shot—and drinks from it absently before setting it aside.
- We repeatedly cut from shots of the cup to shots of the Sebastians, who are studiously looking elsewhere.
- The visiting Dr. Anderson, talking and looking at his interlocutors, reaches down casually to the table where both his and Alicia's cups have come to rest, paying little enough attention that he takes the wrong cup.
- Both Sebastians reveal that they have been attending closely all along by simultaneously and immediately rushing to stop him from drinking from Alicia's cup, finally precipitating her epiphany.

So-called “continuity editing” follows a suite of conventions designed to direct and organize the viewer's attentional experience so as to create the impression of narrative, temporal, and spatial continuity across cuts.¹ These rules of thumb emerged in the first two decades of the twentieth century and were consolidated and standardized in conjunction with the rise of the Hollywood studio system in the 1920s. Today the continuity system is a major component of the normative “Hollywood style,” and the dominant approach to editing in narrative cinema.

Spatial continuity is of primary concern to the continuity system, and many of the rules of continuity editing, including those which we take to be central in the construction of narratively compelling joint attentional scenes, are first and foremost taken as tools for cuing contiguous spaces. Perhaps the best-known principle of continuity editing is the imperative to adhere to the 180° rule, in which the action of a given scene takes place along an axis that establishes a half-circle area in which the camera can be placed. In other words, the camera can take any number of vantages on the scene being presented as long as it stays on the same side of the scene. This rule ensures that relative positions, lines of sight, and direction of movement remain consistent from shot to shot.

Establishing the 180° axis provides an important baseline for the presentation of the attentional dynamics that interest us here, but it doesn't itself rely on social

¹ For a detailed history of the emergence and standardization of the continuity style, see Bordwell, Staiger, & Thompson 1985.

cognition to work. However, the bulk of devices for establishing the relative spatial arrangement of elements from shot to shot in the continuity system are organized around characters' visual attention, dependent on viewers' special interest in and ability to attribute significance to the direction and focus of others' gaze.

The coffee cup scene in *Notorious* makes prodigious and effective use of three of these methods in particular: selectively framed tracking shots that establish the significance of the fatal object. the eyeline match, and shot/reverse shot sequences. The experience of suspense is bred by the deployment of these techniques to set up a progression of unconsummated attentional triangles.

Tracking

Like the opening of *Touch of Evil*, the coffee cup scene in *Notorious* makes use of establishing tracking shots. Here, however, as seen in shots N1, N2, and N3, the tracking focus is not on a character or vehicle traveling through a large space under its own motive force, but on the fatal cup as it travels from Anna Sebastian's hand to the table in front of Alicia.



Figure 5: Shot N1



Figure 6: Shot N2



Figure 7: Shot N3

Selective framing reinforces the overwhelming salience of the coffee cup and its status as a vessel of suspense. It looms in the foreground of every shot of Alicia until she drinks from it. While focus is on her face, the cup always intervenes, often filling

as much as a quarter of the screen. The triangle of filmmaker-spectator-object is unambiguous: the cup is assuredly at the apex of our shared attentions, its significance and prominence inescapable. For the characters, however, the attentional circumstances are quite different.

Eyeline match

In an *eyeline match* edit, one shot presents a character looking at something outside the frame of the shot, and the cut reveals the seen space, sometimes but not necessarily from the exact vantage of the viewing character.

In *Notorious*, the cut from Alexander Sebastian (N4) to Alicia, Dr. Anderson (Reinhold Schünzel), and a coffee cup (N5) is an eyeline match. In the first shot, Sebastian glances up from his newspaper to look at a space offscreen. The second shot shows the space where his gaze was directed, though not from the angle from which he would have seen it.



Figure 8: Shot N4



Figure 9: Shot N5

Bordwell, Staiger, and Thompson (1985: 207-208) emphasize the historical importance of this device in the construction of complex contiguous spatial relationships. The eyeline match is a step forward in subtlety and sophistication from the earliest point of view shots, in which the relationship between contiguous spaces was often indicated by a shot in which a character looks through a distinctively shaped opening (a keyhole, a telescope) followed by a masked shot in the shape of the opening.

Shot/reverse shot

The shot/reverse shot schema coordinates not just the gaze of one character with the object of that gaze, but with a matching eyeline on the other side of the cut. This technique also dates from early in the twentieth century, making some scattered appearances in the 1910s and becoming widely prevalent in the 1920s.² Once the relative position of figures in a scene and the 180° line have been established, the camera can alternate between end points of the line, motivated by the direction of

² See Salt 1992 and Spadoni 1999.

characters' gaze. In the standard, and highly prevalent, shot/reverse shot edit sequence, shots alternate between views of two characters involved in some interaction with one another, usually face to face.

Again, as with the eyeline match, the camera angle is not necessarily aligned with a character's vantage. Often, the camera is positioned visibly behind the rear-facing character, whose shoulder may frequently be visible in the shot. At other times, the rear-facing character's back is not visible, but the framing of the shot suggests that it is just out of sight. In both cases, the camera is often slightly to one side of the rear-facing character's original position; it is rare for the forward-facing character to look directly into the camera in these shots. The camera is thus *aligned with*, rather than *taking the position of*, the viewing/viewed character. Nonetheless, the motivating logic connecting these shots is tightly connected to the gaze and attention of the characters on screen.

As Bordwell (1985: 8) observes:

The filmmaker has this [schema] ready to hand for representing any two figures, groups, or objects within the same place... Because of the tradition behind the schema, the viewer in turn expects to see the shot/reverse shot figure, especially if the first shot of the combination appears... the style's norms not only impose their logic upon the material but also elicit particular activities from the viewer.

The shot/reverse shot schema can be considered a special case of the eyeline match, and the state of affairs Bordwell describes holds for both. Indeed—and this is why we are making such a point of the centrality of these schemas in the highly conventionalized continuity system—all the standard devices of continuity editing are endowed with the logic of joint attention.

A number of edits in this sequence manipulate the eyeline match and shot/reverse schemas to present the viewer with information about where a character is relevantly, sometimes resolutely, *not* looking. The result of this clever reversal of the conventional content of such sequences is an increasingly urgent attentional situation for the viewer, in which the closing of the attentional triangle is as apparently imminent as it can possibly be, but its proper realization is repeatedly deferred. In the final seconds of the scene, all the points are connected in a flurry of activity: Alex and Anna Sebastian both switch from covert to overt attention to the cup; Alicia responds by staring at the cup in dawning awareness; and finally she looks with equal alarm at the Sebastians. All of these moments are conveyed via shot/reverse shot edits, now at last serving their conventional purpose. The interaction of Presentation, Ground, and Reference in the resulting interpretive blend is illustrated in Figure 10.

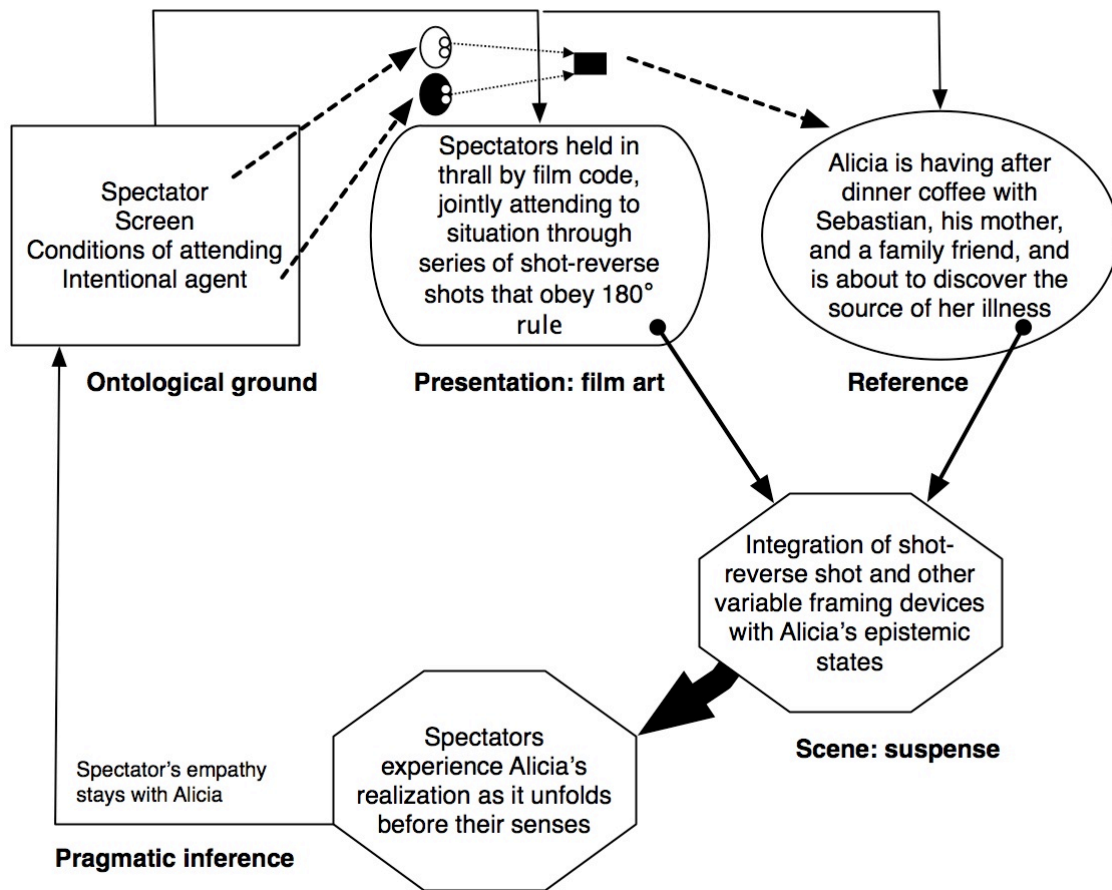


Figure 10: Blending network for the climax of the coffee-cup sequence in *Notorious*

5. Wavelength: Suspense outside the narrative framework

Our principal contention is that suspense in film is an emergent property of conceptual integration according to a standard mental format of joint attention, intention reading, and imitation. Techniques of shot depth, editing, and movement comprise an historically evolved and cognitively motivated palette of "codes" for closing this triangle of attention, with suspense arising as a viewer's response to the "seamless" integration of technique and story, or the scenes and scenarios representing the camera's intentional behavior (i.e., "it zooms in on the man holding an object because it wants me to see that it is a bomb and that the egg-timer is being set for three minutes"; "it cuts to this vantage point because it wants me to focus attention on the cup of coffee that Sebastian spies from across the room," and so on) that characterizes Presentation space, merged with the raw order of events characterized by the Reference space.

Indeed, the classic Hollywood films discussed above provide exceptional examples of this seamless integration, and one can safely assume, with Bordwell and others, that these techniques arouse precisely because they proved to be effective conventions for directing spectator attention on precisely those elements that will suspense and thereby implicate the spectator in a joint-attentional scene that forms the very basis for complex symbolic behavior and communication.

But suspense can also be produced by the "seamed" features or "discontinuities" between Presentation and Reference spaces in minds habituated to integrate film technique and human scale events, or events that fit our cultural models of what is interesting for other people to watch and witness. The genre of "structural film" pioneered in the late 1960's and early 1970's restricts the palette of camera techniques to lay bare the elements of filmic enunciation, so to speak. Michael Snow's *Wavelength* (1967) is a prime example of how a fixed-frame long shot (approximately 45 minutes) can produce its own tensions from many of the same ingredients we have seen in narrative suspense. Here the spectator relies entirely on the camera for closing the triangle.

Comprising no narrative action, *Wavelength's* 45-minute long zoom shot includes four human scale events. The first event is of three people, a woman and two laborers, who move a bookshelf into the apartment, then leave. In the second event, two women chat as "Strawberry Fields Forever" plays on a transistor radio, then leave. The third event is of a man who enters the apartment, collapses on the floor and dies (shot W1). In the final event, the female resident of the apartment is speaking calmly on the phone about a dead stranger in her apartment, whose body now lies off camera (W2). Cinematically, each event corresponds to (1) a different sound wavelength, from low to high frequency, and (2) a different setting on the lens aperture, such that the bank of windows becomes closer and closer, leaving more square footage of the apartment outside the 180° viewing range. Gradually, the spectator begins to figure out that the camera's real object of interest is a photograph hanging on the wall between two of the four windows.



Figure 11: Shot W1



Figure 12: Shot W2

As the camera zooms in and the frequency of the hum increases, the spectator is cued to anticipate that something important is about to be revealed. The content of the picture on the wall will reveal to us the purpose of the film, will reveal the long anticipated secret that makes sense of all these happenings and will reveal what is so important to turn its attention away from such intrinsically interesting and dramatic events as death and dying. We expect this revelation because we know that increasing either the volume or intensity of background music typically foreshadows the presentation of dramatic human-scale actions and events. In the end, however, the viewer learns that the object of the camera's own intentionality is a photograph of waves, nothing else (shot W3). The film becomes a story about itself.



Figure 13: Shot W3

At no point are viewers permitted a seamless integration of human scale events and cinematic technique. They are never permitted to look through the camera at the world on stage. Instead they are relentlessly confronted with the camera as their only means of access to this world, and it is the relentless mechanical regularity of the zoom and the hum that ultimately takes up all our attention. The camera's insouciance toward human scale events of intrinsic interest defines the nature of engagement with the film itself.

The mental format of this cinematic engagement is essentially the same as the previous films but with the distribution of attention weighted heavily on the Presentation space and the camera eye. Instead of a seamless integration of story world and shot sequence, such that camera technique plays a subsidiary, supportive role in the object of suspense, a seamed and "dis-integration" of events and techniques for their disclosure.

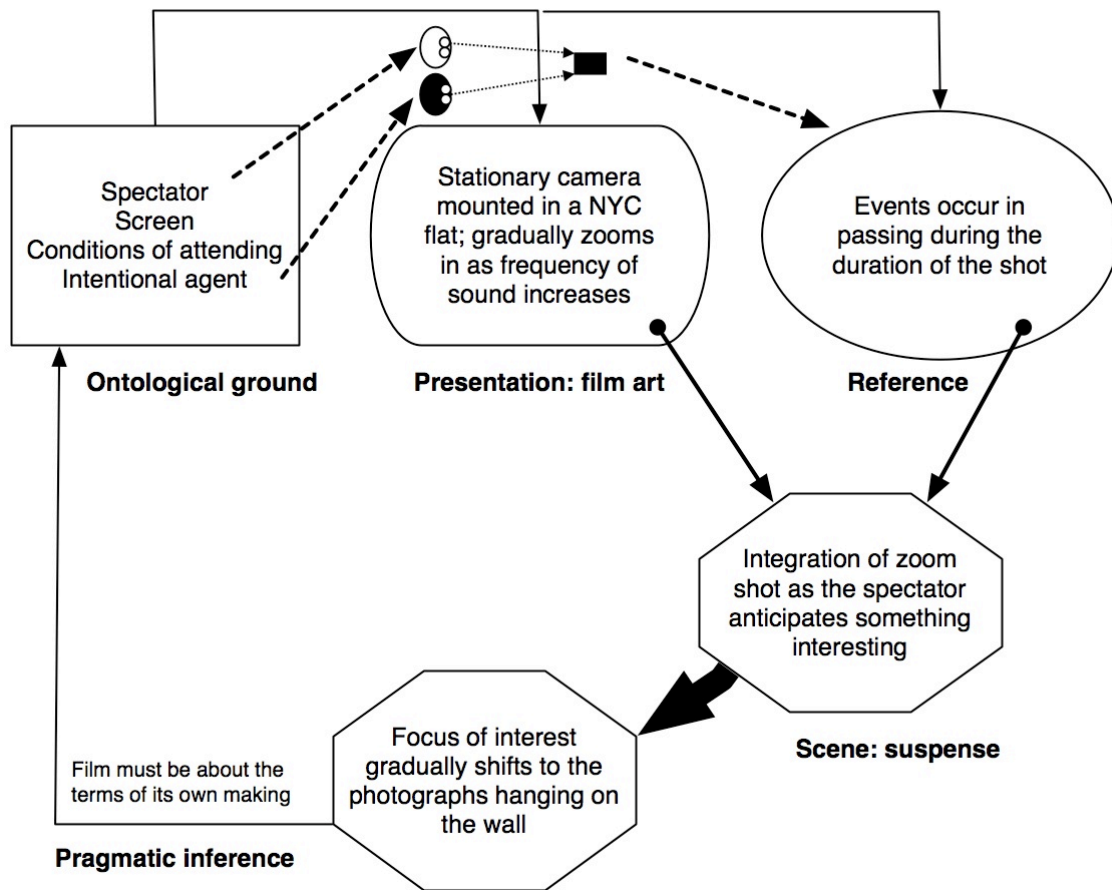


Figure 13: Blending diagram for *Wavelength*

Suspense, then, is not just a product of story worlds. It is an emergent product of our mental format for triangulating attention to common objects of attention. If we may, we would like to refer to *Wavelength* as a suspense film, where the suspense lies in the disjointed relationship between diegesis and the means of presenting it. What spectators are left with and what they experience is a disintegrated integration, where the integration captures the cognitive affective result of being suspended and frustrated, suspended in the sense of having to attach significance to the camera's gaze, with the requisite anxiety or anticipation at finally "closing" the triangle of attention and revealing the "truth," frustrated in the sense of having to attach significance to human-scale insignificant events at the expense of closing the triangle of attention around manifestly significant events.

In short, spectators have a preferred set of categories of interest that this camera eye does not share yet they are powerless to guide it. This is the key to cinematic suspense. The ability to close the triangle is completely the domain and provenance of the camera and filmmaker. No symmetry of co-direction immanent in the basic shared attentional scene is permitted, which is perhaps the major reason why film

has become such a successful method of generating suspense, for suspense is built into the very expectations of filmic engagement, so much so that even the promise of something interesting, even at the negation of its fulfillment, will produce suspense-like states.

6. Conclusion

Suspense requires mastery of one agent over the attention of the other. In the films discussed in these pages, suspense occurs when spectators allow the camera to serve the analogous role of an adult caretaker in a joint attentional scene, whereby the spectator experiences the camera eye as an intentional agent that controls the flow of information. The three features of joint attentional scenes--attention to a third "object," intentionality, and imitative reversal--are imperfectly and inflexibly deployed in the filmic version of interaction, for the spectator cannot influence through imitative reversal the camera's behavior in real time, as human partners in joint activities normally do: the spectator is at its mercy.

Suspense often comes from the spectator's desire to close the triangle of attention, be it to intervene on behalf of the object or to figure out a crucial piece of information, all the while prohibited from doing so according to his or her own preferred temporal window and remaining suspended in a state between anticipation and dread. We think suspense is a common cognitive disposition of film spectators, regardless of genre, and the mental format of comprising two mental spaces more or less seamlessly integrated into a third space provides one plausible cognitive model of how these joint activity between spectator and filmmaker unfolds in real time.

If truly explanatory, however, our joint attentional explanation of suspense should also be able to provide a plausible account of the failure to produce suspense. Enter the so-called "B" movie. A film usually earns this grade when the camera eye no longer masters the spectator's attention, because there is no seamless integration of story world and camera technique, either due to infelicities with one or the other or both.

Consider one of the more famous B movies, Ed Wood's *Plan 9 From Outer Space* (1958), considered one of the greatest examples of "camp" ever produced. A two-time winner of the Golden Turkey award for worst film ever made, Wood's flick chronicles the evil alien's ninth attempt to take over Earth, this time by resurrecting dead. In contrast to the seamless eye-line match and continuity editing of Hitchcock, Wood uses the techniques of continuity editing without much actual concern for diegetic consistency. For instance, at about 6:20 into the film, two sextons from the local cemetery are finishing their task of burying the "old man's" (Bela Lagosi) wife (played by Vampira). As they walk away, they stop and look straight ahead into the distance, disturbed by the noise of a space ship landing (shot P1). The camera cuts to index their line of sight. While the sextons appear in daylight, the camera's cut shows a graveyard at midnight, with an ample dose of fog. From the darkness

emerges the dead wife, zombie-like with long fingernails (P2). Death screams are heard off camera, presumably from the two sextons, off in the distance.



Figure 15: Shot P1



Figure 16: Shot P2

The disparity in time and distance render the causal relation dubious at best, and hence mitigating the intended effect of mortal danger. The public never really becomes fascinated with anticipation and dread, as Hitchcock would have it. Rather, they are more likely to remain focused on the inability of the camera to guide our attention in the intended manner. Suspense then morphs into camp—the love of artifice and exaggeration (Sontag 1964)—with the focus of attention being on the mismatch between story and camera technique. Spectators do not “live in the blend” but live in the interstices of mental space network, but not in the manner exhibited in *Wavelength*, for the camera eye seems to exhibit no intentional will at all. If there is anything like suspense, it arises from anticipation of what the incompetent filmmaker will present next.

Suspense, then, is a function of human social cognition, with joint attention and conceptual integration as the two cognitive processes that give life to the anticipation and dread that accompany our engagement with manipulated story worlds. The cinema has developed an impressive array of techniques for exploiting these features of social cognition.

Bibliography

- Bakeman, R. and Adamson, L. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278-1289.
- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- Bordwell, D. (1985). *Film art: An introduction*. New York: McGraw-Hill.
- Bordwell, D, Staiger, J. and Thompson, K. (1985). *The classic Hollywood cinema: Film art & mode of production to 1960*. New York: Columbia University Press.

- Brewer, W. (1996). The nature of narrative suspense and the problem of rereading. In P. Vorderer, H. J. Wulff, and M. Friedrichsen (Eds.), *Suspense: Conceptualizations, theoretical analyses, and empirical explorations* (pp. 107-127). Mahwah: Lawrence Erlbaum Associates.
- Bruner, J. (1983). *Child's talk: Learning to use language*. New York: Norton.
- Carroll, N. (2001). *Beyond aesthetics*. Cambridge: Cambridge University Press.
- Carroll, N. (2003). *Engaging the moving image*. New Haven: Yale University Press.
- Clark, H. (1996). *Using language*. New York: Cambridge University Press.
- Edelman, G. (1989). *The remembered present: A biological theory of consciousness*. New York: Basic Books.
- Fauconnier, G. and Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.
- Gerrig, R. (1989). Suspense in the absence of uncertainty. *Journal of Memory and Language*, 28, 633-648.
- Hay, D. F. (1979). Cooperative interactions and sharing between very young children and their parents. *Developmental Psychology*, 15, 647-655.
- Hitchcock, A. (1946). *Notorious*. RKO Pictures.
- Oakley, T. (2009). *From attention to meaning: Explorations in semiotics, linguistics, and rhetoric*. European Semiotics Series, Volume 8. Bern: Lang verlag.
- Premack, D. G. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.
- Salt, B. (1992). *Film style and technology: History and analysis*. 2nd edition. London: Starword.
- Snow, M. (1967). *Wavelength*. Joyce Wieland & Michael Snow, Ltd.
- Sontag, S. (1964). Notes on camp. In *Against Interpretation and Other Essays*. New York: Farrar, Straus & Giroux.
- Spadoni, R. (1999). The figure seen from the rear, Vitagraph, and the development of shot/reverse shot. *Film History*, 11(3), 319-341.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. and Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 12, 197-211.
- Truffaut, F. (1967 [1983]). *Hitchcock*. Revised edition. New York: Simon & Schuster.
- Wells, O. (1958). *Touch of evil*. RKO Pictures.
- Wood, E. (1958). *Plan 9 from outer space*. Corinth Films.
- Yanal, R. (1996). The paradox of suspense. *British Journal of Aesthetics*, 36(2), 146-158.