cognitive theory

and the individual film

the case of Rear Window

william seeley and noël carroll

introduction

Motion picture theory—or the theory of the moving image, as we prefer to call it—has been highly suspect amongst movie makers and critics alike. The major source of the skepticism here is the same in both instances. Movie makers and critics care about individual motion pictures—the one they are making or the one they are analyzing. Theory is, well, too theoretical; they want something practical—that is, something that they can put into practice. It is all well and good to speculate abstractly about the nature of the interaction between film and its audience. But the devil is, here as everywhere else, in the details; in the nuts and bolts of motion picture production and in the ways actual consumers are engaged by actual movies. Theory, it is charged, is too broad, too abstract to be of use; movie makers and critics want something more down to earth.

These misgivings are voiced whenever theory is mentioned in the presence of practitioners. And in some cases, such as that of much recent
poststructuralist theorizing, these anxieties are well warranted. However, we would argue that the very best moving image theorizing is not remote from either the production or the analysis of individual movies. Rather, it grows out of these practices in a way that illuminates them. The writings of Rudolf Arnheim, Sergei Eisenstein, and André Bazin, among others, were (and still are) eminently exemplary in this regard.

Cognitivism is a recent entry into the conversation of moving image theory. It only joined the discussion among film theorists in the 1980s, gradually gathering volume over the succeeding two and a half decades. Cognitivist theorizing is apt to strike the movie maker and the critic as something of dubious value, particularly given its roots in the academy. Like previous theorizing, practitioners will question its value with respect to the illumination of individual movies. The purpose of this essay is to allay those doubts. To that end, we will apply the resources of a particular cognitivist framework to Alfred Hitchcock’s classic film Rear Window. In articulating the way Rear Window works from a cognitivist perspective we hope to demonstrate that the approach has broad applicability to the category of mainstream narrative films to which the movie belongs.

Rear Window: an overview

Rear Window, as most readers already know, is the story of a professional photographer L. B. “Jeff” Jefferies, who has been recently injured while covering an auto race. He is convalescing in his apartment in Greenwich Village. He is immobile. His leg is in a cast. He is visited regularly by an insurance company nurse who doubles as his physical therapist (Stella) and by his girlfriend Lisa Fremont. Jeff spends his day at his window—a rear window overlooking an enclosed courtyard—doing what he does best, visually cataloging the unfolding narrative within his purview in terms of the kinds of dramatized themes of human interest that are attractive to newspaper publishers and magazine editors. As the camera pans across the adjoining apartments, the image outside his window resembles nothing so much as a photographer’s contact sheet. Jeff’s attention is particularly drawn to the lives of a traveling salesman and his invalid wife—the Thorwalds—who live directly across the way. His close inspection of snapshot glimpses of their lives leads him to the conclusion that the husband, Lars Thorwald, has murdered his wife, and he sets about proving this hypothesis with the aid of his girlfriend and the therapist.

Given Jeff’s immobility, Rear Window is often interpreted as an allegory for sedentary film viewing. His bay window can be taken to represent the surface of a screen, a window onto a world that is itself presented through and across other apartment windows in bracketed glimpses that resemble cinematic shots and sequences. But the argument can also be made that it is more
aptly interpreted as an allegory of active movie making. Jeff’s eyesight is good enough to provide him with a long view of his neighbors’ doings. Nonetheless, armed with a pair of binoculars and the kind of telephoto lens that a photographer keeps in his toolkit, he can move in for closer and closer views, as he effectively sorts, catalogs, and stitches together a photojournalist’s narrative account of what he observes from a distance. Jeff builds his case against Thorwald—tells his story—by shifting his viewpoint from far views to close-ups—operating as if he were a stand-in for Hitchcock himself, editing *Rear Window* on the fly from within the film out of bracketed glimpses of the lives of his neighbors. Consequently, one may construe *Rear Window* as a self-conscious model for an entire class of movie making and thus a perfect opportunity to display the explanatory power of cognitivism, both for movie makers who are interested in what will work and critics interested in what has worked.

**explaining how *Rear Window* works:** variable framing and erotetic narration

Perceptual psychology has established that our visual system is calibrated to track change in our environment, especially the kinds of relatively abrupt changes that signaled the potential presence of predators or prey for our prehistoric forebears. Movies are typically constructed from a multiplicity of different images, usually in the form of discrete shots. These shot transitions glue the spectator to the screen by titillating this natural perceptual disposition to track and interpret change, thereby securing the movie-maker’s first objective: commanding the audience’s attention. *Rear Window* consists of 816 shots with an average shot length (ASL) of 8.62 seconds distributed over the 110 minutes and 32 seconds of the movie. These figures resemble other Hollywood movies in the 1950s and early 1960s whose ASLs were in the range of 8–11 seconds. This pattern of perceptual change constantly rejuvenates the viewer’s attention to the screen. At minimum, therefore, the very fact that the camera’s shifting viewpoint is finely tuned to the activation of the audience’s innate processes of attention represents an important way in which the movie maker—in this case, Hitchcock—exploits the natural tendencies of everyday perception.

It is not enough, though, for a movie maker to hold onto the audience’s attention through cuts and other means. The audience’s attention must be guided. And if the motion picture in question is a narrative, the audience’s attention must be guided toward what is pertinent for following the story as it unfolds. To this end, movie makers, such as Hitchcock in *Rear Window*, employ a suite of formal-compositional devices in order to assure that viewers attend to precisely what they need to be attending to at any given moment in order to follow the narrative—not only to
continually refresh the viewer’s angle on the action and, thereby, to keep them from turning away from the screen, but also to supply them with the information they need to comprehend the story on a need-to-know basis.

Of these devices—which, like films themselves, we will refer to as attentional engines—variable framing is fundamental.

**variable framing**

*Variable framing* is a mechanism for changing the viewing position—which for convenience we will call the camera position—on the emerging course of events and/or states of affairs in the movie world. Movie makers use variable framing to direct attention to salient information by changing the viewer’s visual perspective on depicted objects, states of affairs, events, and actions. A striking example of variable framing occurs in the opening of *Rear Window*. The camera moves from a shot of Jeff’s leg in a cast, to a broken camera on a side table, to a photograph of a car crash above it on the wall, two race cars looming in the foreground, tumbling towards the viewpoint of the photographer’s lens. By selecting these details and framing them saliently—that is, by guiding the audience’s attention to them and controlling the context of their presentation—Hitchcock deftly and economically explains how Jeff broke his leg. The camera then moves on to draw our attention to a photograph of an atomic explosion, the negative of a fashion model’s head-shot, and finally a magazine cover featuring the same photograph. Thus, by means of variable framing, Hitchcock informs us that Jeff is a magazine photo-journalist and foregrounds our introduction to his girlfriend, Lisa Fremont, suggesting that they met on a fashion shoot when we meet her in a later scene. This sequence of selective reframing offers an illustrative lesson in how movies narrate visually. It also supplies a key for the viewer that indicates how variable framing will be used to structure the rest of *Rear Window*.

There are three prominent ways to implement variable framing: editing (cutting from one camera position to another); camera movement (panning or tracking from point A to point B); or lens movement (zooming in or zooming out). These physical means of altering our viewpoint on the action, furthermore, illustrate and support the three basic functions of variable framing: *indexing*, *scaling*, and *bracketing*. Indexing refers to the rudimentary capacity of the camera to call our attention to something by pointing at it. When the camera moves toward something by cutting, tracking, panning, or zooming in, it functions ostensively, pointing towards some feature of interest in the scene. Its direction signals “Look at this!” Ostension or simple pointing is perhaps the most basic form of human communication. When the camera shifts its position on the action in this way, it typically indexes something that plays a critical, contributing role.
in the viewer’s capacity to recognize and understand the unfolding of the story. For example, when we see a close-up of Jeff’s thermometer, this informs us of the stifling heat that, in turn, establishes an insistent source of irritation, reinforcing the atmosphere of edginess that suffuses the story world.

Scaling is typically a natural consequence of indexing in contexts like these. As the camera moves closer to the target subject of the shot, it takes up more space in the visual field. The attendant change in the size of the target indicates its importance to the story. Changing the relative size of an object or agent on the screen is a simple but important way of communicating its importance. As a general rule of thumb, what is large on screen is important; conversely, what is small is of lesser significance. Of course, there are cases of indexing without scaling—that is, without altering screen size. Relative movement can achieve the same effect by detaching an object or agent from the background in a scene—for example, tracking shots which maintain the subject in a central position against a moving and changing backdrop, or a pan of a character as he or she walks a short distance. The opening camera movement in Rear Window, which introduces us to the courtyard, is an instance of indexing without scaling. It is akin to a sweeping gesture—indicating in effect that this whole environment is to be the locus of our attention—without changing the relative scale of the apartments. However, scaling and indexing most frequently occur in tandem insofar as typically shifting indices often automatically involve alterations in scale.

A third function of variable framing is bracketing. Like scaling, it is also generally a natural consequence of indexing. As the camera viewpoint is moved by editing, tracking, panning, or zooming towards the subject, elements of the scene once visibly present become occluded beyond the frame of the screen, altering the context of presentation. The frame acts as a bracket. What is outside the frame has been literally bracketed out and is no longer available to be seen. What lies within the bracket onscreen is most often what the viewer needs to attend to. When Lisa signals to Jeff that she has Mrs. Thorwald’s marriage ring, the close-up not only points to what is significant and marks it by enlarging the amount of screen space it occupies, but the shot also excludes any potential distractions. Moreover, this compositional strategy can also be used to alert us to something off screen, as when a character looks apprehensively beyond the frame of the shot, when a noise penetrates the scene from outside the bracket, or when the viewer has been made aware of, or been made to expect, something offscreen that the characters are either unaware of or unprepared for. For instance, when Lisa suddenly looks apprehensively off screen and encourages Jeff to recount his suspicions about Lars Thorwald, we are alerted to the fact that the next important item of narrative information is outside the frame—here, the fact that Thorwald is binding a large piece
of luggage with rope. This kind of exclusive (as opposed to inclusive) bracketing demands active perceptual and cognitive engagement. It is incumbent on a viewer to track, update, and interpret the indexed information off screen by tracking and interpreting the actions that are explicitly perceptually present in the scene.

**erotic narrative structure**

Variable framing can be used for a range of purposes in the movies—for example, to depict an environment; to draw contrasts and comparisons among actions, events, characters, and settings; to make and support arguments; or to express emotions. However, we believe that there can be little doubt that the most common and paradigmatic role of variable framing is to develop and articulate movie narratives by guiding attention to critical story information, highlighting the salience of this information via indexing, scaling, and bracketing, and presenting it in a sequence that facilitates our construction of a coherent, intelligible, unified, and compelling story. This is a process that requires active cognitive engagement on the part of viewers. Moreover, it suggests that cognitivist theory is a resource that can be fruitfully employed to track and explain the range of ways these devices have been used to engage key aspects of human cognitive and perceptual architecture in order to articulate the particular narrative structure of particular films. In the remainder of this section we will explore the use of variable framing in concert with the erotic narrative structure of Hitchcock’s *Rear Window* in order to motivate this hypothesis.

Movies are constructed from sets of spatio-temporally disjoint shots, scenes, and sequences that adeptly direct viewer attention, in order to deliver just what information is needed just when it is needed for the development of a globally-cohesive overarching narrative. This formal-compositional fact about movies raises two questions: What binds spatio-temporally disjoint shots and sequences into locally coherent scenes, actions, and events? What are the governing principles viewers use to stitch these spatio-temporally disjoint scenes, action, and events into a unified and cohesive narrative whole? We argue that the answer to both of these questions is that most movies generally employ erotic narrative structures.

Erotetic narratives are constructed by generating and answering questions about the actions and events depicted in a text or movie and then answering them. In ordinary run-of-the-mill contexts, audiences respond to depicted actions and events by asking questions about their causes and effects—about where they came from and where they are headed. Actions, the goal-directed intentional behaviors of characters, are the primary subject matter of the movies. Actions differ from events in that their causal source lies primarily in the mental states of the acting agent. Our inferences to causes in answering erotic narrative questions at the movies will
therefore ordinarily involve interpretations of the beliefs, desires, emotions, and motives of characters as psychological causes of target actions. The fragmented presentation of information in the scenes and sequences in movies thereby prompts questions about the causes and potential effects of depicted actions and events that the movie goes on to answer. The attribution of psychological states to characters also serves two further purposes in these contexts. It opens up a range of possibilities, raising questions about what will happen subsequently in the narrative—how the characters will behave, what they will do, and how they will interact with one another. It can also bring to light information about foregrounded objects, events, and actions, retrospectively framing their salience in the context of present narrative possibilities. The attribution of mental states to characters therefore guides what we call the forward and retrospective articulation of narrative structure.

Consider for instance a scene that plays out across our bracketed views of the Thorwald’s apartment early on in *Rear Window*. It is the middle of the morning. Mr. Thorwald has just returned home with his samples case. He takes off his hat, jacket, and tie, and goes into the next room where his wife is in bed. They argue. Mrs. Thorwald points to her watch, narrowing our focus within the bracket of the bedroom window while she admonishes him about the time. We infer that she is nagging him, asking, “Where have you been all this time?” This automatically leads us to wonder where this marriage is headed. That is, by bracketing the couple’s argument and indexing the current time, audience members are alerted to be on the lookout for new information, the assimilation of which will contribute to their understanding of these local and global narrative questions within the Thorwald’s sub-narrative narrative, thereby enabling them to accommodate it and retroactively update their grasp of the story.

Variable framing is generally coordinated with erotetic narration, as this example of indexing (without scaling) and bracketing suggests. In the typical case, variable framing functions to assure that the first item or gestalt of items the viewers see with a new camera viewpoint is the item or items that are most relevant to following the evolution of the story at that point in its presentation. This may involve the posing or answering of questions, or both. In the current example, our attention is focused on the odd timing of Mr. Thorwald’s return home, an action that is initially framed for the audience by a view of his wife’s aggravated but inquisitive posture in the next room; this view is inclusively bracketed for the viewer by the bedroom window, thus serving the erotetic narration by raising questions (arousing curiosity) about the status of this relationship. In sum, variable framing is used in order to ensure that those elements of a cinematic sequence are salient that will advance the network of questions and possibilities that drive the narration forward.
There are two major classes of questions that movies pose: micro-questions and macro-questions. Micro-questions emerge within scenes and sequences and supply the connectives between shifting camera positions. For instance, Lisa's face becomes worried and inquisitive, raising the question of why, and then the variable framing answers that question immediately by foregrounding Thorwald tying up the large trunk. Here, once again, the variable framing drives the local erotetic structure of the sequence.

Macro-questions are overarching questions that sustain our attention through all or most of the story. In *Rear Window*, two such macro-questions drive two, hierarchically related sub-narratives: a macro-question about the evolving status of Lisa and Jeff's relationship drives the global romance narrative and a subordinate macro-question about the outcome of Jeff and Lisa's sleuthing drives the dynamics of their relationship. The latter macro-question is itself articulated by three subordinate narrative questions: will Jeff be able to convince Lisa of Thorwald's crime; once he does will they together be able to collect enough evidence to convince the police of Thorwald's guilt; and once they do will Thorwald thwart them in their effort, bringing the romance narrative to a close with their tragic demise?

The various subplots that unfold within the frames of the neighbors' apartment windows are likewise driven by their own subordinate macro-questions, each of which itself articulates a question salient to one or both of the romance and adventure narratives. Will Miss Lonelyhearts find a mate? Will the composer successfully finish his new song? Who will Miss Torso choose? These questions each evoke their own narrative possibilities. As the story unfolds the range of available possibilities is pruned, until finally we have answers to all of these pressing questions. Thorwald is captured, Lisa and Jeff are closer than ever before, Miss Lonelyhearts is romancing the successful composer, and Miss Torso's true love has returned from the wars. Thus the individual narrative questions have been answered, resolved, or explicitly set aside for a sequel, and the narrative comes to an end, or achieves closure.

**erotetic narrative and variable framing**

Lens movements, camera movements, and cuts function to variably frame narrative elements by means of indexing, scaling, and bracketing. These activities are used to visually assert the significance of information within shots and cinematic sequences, thereby guiding attention and shaping the narrative. As we have seen, indexing can be achieved by moving the camera viewpoint closer to the subject, scaling the visual perspective on the target. As the camera position moves in closer to the subject it increases in visual size, indicating its importance. Consider the initial sequence with the telephoto lens. It begins with a long shot of Jeff's unassisted perspective on
Thorwald cleaning his samples case. His piques his curiosity. He asks Stella to retrieve his binoculars. Now his perspective is moved closer in. The action is scaled larger, from a long to a medium shot that reveals Thorwald restocking his samples case. When Jeff changes to his telephoto lens, he perceives Thorwald cleaning his saws and wrapping them in newspaper in the kitchen; this action is scaled larger again, in a close-up, signaling that this information is even more visually important. Here indexing functions as a means to collect evidence in support of Jeff’s hypothesis about Thorwald’s crime, which prunes the range of narrative possibilities and narrows the range of possible answers to the question of what has happened to Thorwald’s wife.

Hitchcock exploits scaling as an attentional engine throughout the movie to deliver information critical to its erotetic narrative structure. For instance, the thermometer fills the screen in close-up each time it is shown, signaling the importance of the information it carries. The pan across the courtyard in the opening sequence comes to rest on an extreme close-up of Jeff’s supine face and head, which overflow the screen. The camera lingers here momentarily to capture a bead of perspiration break free and roll across his head. As the bead of sweat reaches the hairline above his sideburns the camera cuts to the thermometer, scaled to overflow the screen as well. It reads a sweltering, claustrophobic 94 degrees, signaling the start of an oppressive New York City summer day, a time when tempers are apt to be short, anger ready to flare, and anything might happen. In this sequence, scaling is used to narrow the focus of our attention: it begins with an establishing shot of the setting, next moves to the establishment of Jeff’s importance as the main character, and then captures his confined state of mind, trapped as he is in the hot and stuffy apartment. Scaling visually foregrounds information, the significance of which is amplified by Jeff’s subsequent telephone conversation with his editor, to whom he comments, “If you don’t pull me out of this swamp of boredom I’m going to do something drastic.”

The thermometer plays a similar framing role, at the start of the scene in which Lisa comes to recognize the validity of Jeff’s hypothesis. Here it reads an oppressive 82 degrees at 10 o’clock at night. In this case Hitchcock uses the shot to accentuate the close, frustrated tone of their bickering. Finally, the same close view of the thermometer opens and anchors the closing pan that marks the end of the movie. Here it reads a comfortable 70 degrees, signaling that the heat, like the romantic crisis in their life, has broken. New York has settled into a comfortable temperate climate that mirrors their now secure relationship.

An analogous story can be told about the way variable framing is used to indicate the narrative salience of Jeff’s telephone. It is first introduced as the center of our attention on the side table when the camera zooms out from Jeff’s cast in the opening sequence, prior to panning to the broken
camera, foregrounding his phone dialog with the magazine editor. It is likewise scaled to fill the screen in the shot following Lisa’s realization about Thorwald’s crime. When it rings, it is Lisa calling to report that she has managed to discover Thorwald’s name and apartment number, thus initiating the adventure narrative and her role as Jeff’s physical proxy in action and risk. Finally, it is Jeff’s gaffe on the phone that leads Thorwald to him as the architect of the investigation in the penultimate scene.

Bracketing (inclusive) is a natural consequence of Jeff’s use of the binoculars and telephoto lens to watch the Thorwalds and Miss Lonelyhearts. The oculus used to denote this mode of attention reveals the details of Jeff’s obsessions and signals his myopic focus. This myopic focus also produces exclusive bracketing in some cases. Consider the pan from the alley to Lisa and Stella’s feet and the hole they have dug in the garden. This camera movement calls attention to the failure of the search for evidence here, to the lack of evidence within the frame of the bracket. But it also signals that Jeff has forgotten to keep watch for Thorwald’s return through the alleyway that is now outside the bracket. Similarly, exclusive bracketing is critical to Lisa’s encounter with Thorwald in his apartment. We cannot see where she has gone to hide in between the rooms. Thorwald’s body language not only tells us that he has found her in this bracketed space, but also allows us to interpret her unseen actions in response to his behavior. In each of these cases, narrative expectations about what we currently cannot see outside the bracket of the camera give life and pressing significance to particular narrative possibilities, increasing the level of suspense in the scene.

The range of questions and answers that define erotetic narrative structure can, as discussed above, be arranged hierarchically into micro-questions, macro-questions, and presiding macro-questions. Presiding macro-questions organize the plot as a whole. In *Rear Window* the presiding macro-question is ultimately the question about the outcome of Jeff and Lisa’s relationship negotiations, which structures the romance narrative. Simpler macro-questions are questions that organize large portions of the movie, but do not structure the plot as a whole. The adventure narrative drives a significant part of the second half of the movie. It can be divided into two overlapping subordinate narratives: a mystery story and a suspense story. The mystery story is structured around the question of what has happened to Mrs. Thorwald, and the subsidiary questions of whether Jeff can convince Stella and Lisa that something is awry and whether, once he has convinced them, they can together collect sufficient evidence to convince Detective Doyle of Thorwald’s guilt. The suspense story takes over the structure of the mystery story once Lisa and Stella leave the apartment in search of evidence of Thorwald’s crime. It includes an additional question of whether Thorwald will be apprehended before he harms either Jeff or Lisa.
Micro-questions are questions about local obstacles to the overarching goals of the protagonist defined by macro-questions. The solution to these problems are, more often than not, related to macro-questions as a means to an end. The use of variable framing to pose and answer micro-questions is thereby a critical device that contributes to the resolution of macro-questions. For instance, the suspense in the scene where Lisa searches Thorwald's apartment is driven by the question of whether she will be caught. The local obstacle to her escape is her motivation to find Mrs. Thorwald's handbag and wedding ring. The discovery of her ring is a means to the end of establishing Thorwald's guilt, solving the mystery, establishing closure in the adventure narrative, and thereby securing a positive outcome to the romance narrative. Variable framing is used to raise and then answer these questions, first by indexing the emptiness of the handbag, focusing on a medium close-up of Lisa shaking it out to show Jeff the ring is not there, and then, once the police have arrived to save her, scaling the view of the wedding ring on Lisa's finger in a close shot to show that she has found it.

Answering this micro-question—will she find the ring?—answers the macro-question about what has happened to Mrs. Thorwald. It resolves the mystery narrative, secures the outcome of the suspense narrative by providing the evidence necessary to convince Detective Doyle to apprehend Thorwald, and thereby secures the outcome of the adventure story.

point-of-view editing

Although the moving image is a cultural invention, it is useful never to lose sight of the fact that in many ways it succeeds as well as it does because it freely avails itself of our biological heritage. The motion picture is an art form, but, as with art in general, we must not suppose that it is solely an affair of the mind. Motion pictures address our bodies as well.6

Point-of-view editing (POV) is a powerful attentional strategy in the movies which is particularly fine-tuned to our biological repertoire of cognitive routines. Minimally it involves the deployment of two shots, a point/glance shot and a point/object shot.7 The point/glance shot captures the gaze direction and facial expression of a character looking offscreen. The point/object shot putatively captures what the character sees. These shots can be related both prospectively and retrospectively. Prospectively, the point/glance shot suggests the mode of attention the viewer should adopt by revealing the current emotional disposition of the character. Retrospectively, the content of the point/object shot, in turn, articulates and clarifies the emotional state expressed by the character. The structure and role of these shots can be iterated in a range of ways. The character can be depicted alone or in a crowd. The character can be the central focus of the shot or a tangential element that serves to capture and shift the focus.
of attention. Multiple instances of the two types of shots can be strung together to guide attention through an unfolding event or action.

Our capacity to follow what other people are attending to in the environment by tracking their gaze is a natural mechanism for interpreting their goals, interests, and emotions in ordinary social contexts. POV in cinema builds upon this innate propensity to track a glance to its target, juxtaposing the character’s look to what is looked at, while deleting the space in between. POV enhances the resources of variable framing by adding the gaze of the character to indexing, scaling, and bracketing as another lever for focalizing what is relevant in the story world and directing the audience’s attention to it. Moreover, it is an especially smooth device for accomplishing this, since, as we have noted, it mimics a natural perceptual pathway—that of following the trajectory of another’s glance to its target.

POV is a perceptual strategy that is regularly used to guide and control the viewer’s attention in *Rear Window*. It is the predominant visual strategy in the film. The number of object shots in the film that are covered as being within (or possibly being within) the gaze of another character is quite stunning. Consider the scene where Jeff observes Thorwald cleaning his samples case and knives. The sequence begins with a shot of Jeff; his attention is sharply focused on what is happening across the yard, behind the camera position. A cut to Jeff’s view reveals a long shot in which we can see Thorwald cleaning his sample case through his living-room window. But the view is not very detailed. Jeff asks Stella for his binoculars. We see him bring them to his eyes. The camera cuts to a medium close shot of Thorwald restocking his cleaned sample case. Jeff puts the binoculars down, and with a curious but concerned look rolls back into the shadows. He installs a telephoto lens onto the body of his camera. When he looks again we see a close shot of Thorwald in the kitchen, cleaning some big saws and knives, wrapping them in newspaper, stretching, yawning, and retiring to the couch for a mid-afternoon nap. When the camera returns to Jeff he pulls pensively away from his camera; a look of concerned horror washes over his eyes. The structure of this sequence provides just the information we need to interpret the thoughts and emotions that underwrite Jeff’s facial expressions and link them to their objects. It also guides the viewer, here providing the viewer a key to the use of POV editing throughout the rest of the movie.

One might think of these processes as putting ourselves into Jeff and Lisa’s shoes, as automatizing information transfer in our experience of the movie by hijacking a natural attentional strategy. However, we think this is the wrong way to interpret what is going on. We think, rather, that POV sequences are representations of the visual experiences of characters that we easily recognize and assimilate because they ride piggyback on a structural resemblance to our own natural perceptual practices. In other words, to the degree that we simulate perspectival seeing in our
engagement with point/object shots, we do so in order collect information salient for understanding the depicted visual experiences of characters. However, we do not, in so doing, put ourselves into the narrative. The easiest way to see this is to note that precise eye-line matching is not a necessary condition for the success of POV sequences. A glance in the right general direction will do. In fact, in *Rear Window*, although Jeff’s optical perspective sits at one end of the axis of action, it is often quite difficult to interpret the precise direction of his gaze vis-à-vis the different apartments across the way.

This all makes sense structurally. In ordinary contexts, we do not track someone's gaze literally as a line traced in space. We are rather presented with its endpoints, the glance and its target, which naturally fall at the two ends of a saccade. A POV sequence likewise delivers the glance and target as discrete objects, as the two endpoints of a natural attentional strategy. If eye-line matching was systematically inconsistent or consistently out of line with the spatial structure established for a scene, we would likely be distracted. But as long as there is a general match to direction, everything should work out fine. Viewers recognize POV sequences as what they are, representations of the perceptual experiences of viewers designed to direct our attention to critical information and to articulate the mental states of characters.

Directors use POV sequences as a communicative device to deliver narratively salient information. These sequences perform this function fluidly because their structure is fine-tuned to natural, biologically-rooted attentional strategies that we deploy in ordinary social contexts in order to collect information about the goals, interests, and mental states of others, as well as the target objects of their intentional behaviors. What this should make clear is that the purpose of POV shots is rarely just to deliver information about the target of the point/object shot alone. Their broader purpose is to articulate a relationship between the perceiving agent and the object of their perception, which is constitutive of the mental and emotional states of characters. Facial expressions are usually not neutral. Or perhaps better, even a neutral facial expression is just as informative about the mental states and emotional dispositions of an agent in a context as are its angry, joyous, or fearful cousins. For an example of how the face in the point/glance shot emotively evaluates its object, consider the way in which the anguished look on Jeff’s face imbues the shot of Thorwald wrestling with Lisa with heightened anxiety.

Movie makers exploit this fact to guide the modes of attention viewers bring to bear on the point/target shot in POV sequences. Again, this is a consequence of our natural perceptual repertoire. There is a close coupling between motor mimicry, affective response, and attention in facial recognition. We naturally mimic perceived facial expressions. The low-level muscle responses associated with this process are accompanied by autonomic responses that are congruent with the autonomic responses
constitutive of the experience of the emotions expressed, even in the context of perceived pictures of expressive faces. These autonomic responses are, in turn, the consequence of affective processes that are also responsible for priming perception to the salience of a stimulus to an agent’s well-being. This entails that our affective dispositions actively shape attention and perception. Thus, information present in the point/glance shot prospectively inflects the affective narrative salience of the target before we see it in the point/object shot.

This relationship is reciprocal. The point/glance and point/object shots perform different kinds of functions. The point/object shot, in conjunction with the tools of variable framing, focuses attention on critical pieces of narrative information. The point/glance shot prospectively inflects the presentation of that information by shaping the mode of emotive attention a viewer brings to bear on the point/object shot by ascribing valenced affective salience to its target. The object in the point/object shot, in turn, retrospectively refines our understanding of the emotional states expressed in the point/glance shot by articulating its content.

The Kuleshov Effect—named after the early twentieth-century Russian filmmaker Lev Kuleshov, who designed them—can be used to illustrate how pov sequences are used to articulate the emotional states of characters. Kuleshov cut shots of the expressionless face of an actor with shots of a range of scenes and objects in these experiments, for instance a bowl of soup; a woman on a divan; a girl in a coffin; and flying birds framed against the sky. He reported that viewers perceived the actor’s facial expression differently in each case, taking it to express hunger, desire, grief, and a yearning for freedom respectively. Kuleshov’s demonstration was not a controlled experiment. Nonetheless, it illustrates the communicative power of editing strategies in film. Viewers naturally connect juxtaposed images across cuts, binding them in space and time. pov sequences are perceived in this way as inflections of the mental states of characters, their thoughts, goals, desires, and emotions. Of course, the point/glance shots are not ordinarily built from neutral expressions. Rather, just as persons acting in ordinary contexts, movie actors are depicted with facial expressions that are congruent with their narrative context, with the way they are situated in their environment, with the bearing that the target of their gaze is likely to have on their well-being or that of others. A case in point is the helplessness and anxious fear expressed by Jeff’s face and body language throughout the point/glance shots that depict him watching Lisa search Thorwald’s apartment and get caught.

We readily recognize gross basic categories of emotions like anger, fear, disgust, contempt, joy, sadness, and surprise in expressive faces in ordinary contexts. There is evidence of cross-cultural convergence in this capacity. Individuals across cultures tend to agree in their categorizations of the
basic category of emotions expressed in close-up photographic portraits of
decontextualized expressive faces. This convergence is higher in cases in
which the photograph has been posed to amplify features diagnostic for
the expressed emotion (and de-emphasizes muscle responses associated
with other aspects of ordinary behavior such as speaking, shifting gaze, or
reorienting one’s face to track the dynamics of scene). The cross-cultural
capacity to recognize the emotions expressed in decontextualized posed
portraits suggests that this is also a biologically-rooted, innate aspect of our
basic, shared perceptual repertoire. In film contexts it enables directors to
prospectively inflect the narrative salience of the target of a point/object
shot, and enable viewers to bootstrap their way into an understanding of
characters and the contents of scenes prior to the elaboration of narrative
context, perhaps providing the foundation of narrative understanding.

This account of POV provides an explanation of one mechanism
directors can use to communicate the basic general categories of emotions
expressed by characters. However, emotions are more fine-grained than this.
They are motivational states that encode the value of their object to an
organism in a context. We do not flee fear. We flee because of our fear of
something. As philosophers say, emotions have intentionality. They are
directed at aspects of the world that are constitutive of their content. The
character and motivating force of our fear, for instance, is determined
by the bearing its object has on our well-being and that of others. This
suggests that the emotion expressed in a point/glance shot alone is under-
determined, ambiguous, or vague, a complication that is further exacer-
bated by the fact that it is difficult to disambiguate some basic emotions,
for example, distaste from concern, or fear from shock or surprise, in facial
expressions.

The solution: POV sequences are fine-tuned to the structure of our
ordinary perceptual recognition capacities. Just as we use the target of an
agent’s gaze to disambiguate and articulate the particular content of their
emotional states in ordinary contexts, the point/target shot delivers the
information requisite to retrospectively articulate the emotional states
expressed in the point/glance shot in POV sequences—POV shots are both
prospectively and retrospectively articulated. Consider Lisa’s expression at
the end of the dialog preceding her recognition of Thorwald’s guilt.
She has been testily admonishing Jeff for his putatively unhealthy interest
in the personal lives of his neighbors. She looks towards the drawn shade
of the newlyweds next door and says, “Why, for all you know there may
be something far more sinister going on behind those windows.” The
camera cuts to Jeff, who turns to see what she is looking at, and then he
turns back to the camera and says, laughing, “No comment.” (It has been
insinuated that they have been incessantly making love for days on end.)
The camera cuts back to Lisa. Her expression transforms from mild
amusement to distracted reflection, and then to deepening concern. Given
the current context framed by the immediate sequence of shots and
current dialog, we initially read the distracted reflection as a self-reflective
response to Jeff's sarcastic comment about the newlyweds.

But her look of concern deepens and she slowly stands. The camera cuts
to Jeff. His eyes follow her as she rises off-frame. The camera cuts back to
a medium shot of the two of them. Her gaze has become intently focused
on events outside the window. Jeff retrieves the binoculars. When the
camera cuts back the courtyard, we are shown a medium shot of Thorwald
tying up the trunk, thereby resolving and clarifying the source and
emotional tone of Lisa's concern. Clarifying the focus of Lisa's gaze thus
disambiguates the emotional content of her original facial expression,
retrospectively articulating that her mental state is one of externally-
directed apprehensive concern, not of internally-directed, self-reflective
remorse.

Here we see that the point/glance and the point/object shots interact
reciprocally in promoting our understanding of what is going on. The
point/glance shot, by dint of Lisa's facial expression, gives us a global sense
of her attitude toward what she is seeing; she is somehow concerned. But
her visage does not fully define the nature of that concern. Rather the
point/glance shot does, by showing us what she is concerned about. Her
concern involves apprehensiveness, since the shot adds to her calculation
of the probability that Mrs. Thorwald has been murdered. That is, the
point/glance shot yields a broad characterization of Lisa's emotional state—
it establishes the range of emotions that Lisa might be experiencing. And
then the point/glance specifies the particular emotional state by showing
us or focusing upon what Lisa's concern is about.

The point/object shot works this way insofar as the emotions are
appraisals that are governed by criteria. The various emotions have criteria
for what will count as the appropriate sorts of objects for that kind of
emotion. Thus, we glean the emotional state of movie characters by
matching the emotional facial expression found in the point/glance shots
with what the point/object shots exhibit in terms of the criteria of emotive
appropriateness that they satisfy. The point/object shots can operate in this
way because they have been pre-selected by the movie maker by virtue of
the emotive criteria they fulfill. In this regard, we say that the objects in
the point/object shots have been *critically prefocused*—that is, they have been
selected and staged because they are the sorts of things that are apt to raise
the affects the movie makers wish to evoke at that point in the motion
picture. Apprehensiveness is what Hitchcock is aiming for, so he fills in Lisa's
visual field with a gruesomely suggestive object.

Criterial prefocusing is one of the primary ways in which movie makers
govern the affective reactions of spectators. By means of variable framing,
they select and make salient objects that are criterial for the elicitation of the emotional states they mean us to occupy. The emotions are attentional engines in everyday life and they perform in a comparable fashion in motion pictures—both to rivet our attention to the screen and to mold our affective response to it. In most narrative movies, criterial prefocusing is a leading mechanism for securing audience engagement. Moreover, criterial prefocusing can obtain within or without a point of view structure. However, although criterial prefocusing occurs quite often without POV structures in a great many films, in *Rear Window* it is rarely deployed independently of said structures, since almost every image in Hitchcock’s film is or could be the point/object shot of some or another character (including the composite view of the assorted neighbors, conceived of as a collective). Thus, by employing the cognitivist concept of criteria prefocusing, one is able to zero in on one of the most distinctive and creative dimensions of Hitchcock’s achievement in *Rear Window*, namely the use of criterial prefocusing almost exclusively within POV sequences as a formal-compositional device, as an attentional strategy to direct viewers’ narrative engagement with the movie and its characters and articulate its content. 

**summary**

Theoretical approaches to the motion picture are often criticized for having little or nothing to say about individual movies. Although this blanket reservation might also be leveled at cognitivist approaches, in this chapter we have attempted to defeat this worry by illustrating the way in which a cognitivist approach to Hitchcock’s *Rear Window* can illuminate how that movie works to hold and guide the viewer’s engagement with the film. 

*Ex hypothesi*, this should be of interest to movie makers and analysts alike—to those who wish to make movies by recruiting reliable strategies, and to those who wish to figure out how specific movies manage to shape audience response.

We have introduced a series of cognitivist concepts as well as hypotheses about their applicability. These include the notions of variable framing, indexing, scaling, bracketing, erotetic narration, micro-questions, macro-questions, and criterial prefocusing. These concepts, we maintain, enable movie makers and critics alike to understand how Hitchcock succeeds in engaging his audiences. Moreover, our framework helps to explain the way in which Hitchcock used criterial prefocusing to articulate the content of POV shots, isolating and articulating perhaps the most distinctive stylistic feature of *Rear Window*.

Thus, we hope to have demonstrated that cognitivism need not be rejected as offering at best a superficial account of particular works. Rather, it is arguably a way of understanding movies ever more deeply.
notes
4 Although erotetic narration is not the only analytic model that has been proposed for understanding movie narratives, we think it is a powerful tool that is well-suited to handle the structure of most popular movies.
5 An inference prompted and reinforced in part by the concurrent phone dialog between Jeff and his editor about marriage and relationships.
6 Carroll, *The Philosophy of Motion Pictures*, 189.
7 Edward Branigan, *Point of View in the Cinema*, Approaches to Semiotics 66 (Berlin: Walter de Gruyter, 1984); 103.
10 See David Bordwell and Kristin Thompson, *Film Art: An Introduction*, 8th ed. (New York: McGraw-Hill, 1999), 227–229. Kuleshov’s own claims about the results of his experiments differ from our claims about point/glance shots. He did not think this device contributed to a viewer’s understanding of the shot/sequence. Given that the actor in the shot was utterly expressionless, he thought the viewers had misperceived the sequence.
A “formal/functional” approach has arguably been one of the central threads in cognitive film theory. The approach, broadly construed, assumes that formal elements and structures of film function to elicit specific effects in audience members, and this perspective has been taken in work on, for example, narrative; character predicaments and emotion; and characters and sympathy; to name some of the main areas of study. The formal/functional perspective is adopted in this chapter to show the way in which the formal structures of many video games fit with fundamental structures of human cognition. The chapter thus gives a bare bones sketch of a cognitive poetics of video games: video-game design is, like other mass and popular arts, craftsmanship under specific conditions. An important cross-cultural aspect of these conditions is human cognition, in that cognition both constrains and enables successful video-game design. I will focus on certain video-game genres, mainly action and adventure, and this will gloss over many nuances: the purpose is not to defend general claims about all video games but rather to show the relevance of particular aspects of cognitive theory, which have not yet been at the forefront of