

# SHOWING SEEING IN FILM

SAMUEL CUMMING, GABRIEL GREENBERG, ELSI KAISER  
AND RORY KELLY

In this paper, we describe two film conventions for representing what a character sees: POINT OF VIEW (POV) and SIGHT LINK. On a POV interpretation, the viewpoint of a shot represents the viewpoint of a particular character; while in sight link, a shot of a character looking off-screen is associated with a shot of what they are looking at. Our account of both treats them as spatial in nature, and relates them to similar spatial interpretative principles that generalize beyond character eyelines. We additionally observe that whether the object shot in a sight link takes a POV interpretation depends on whether the glance shot precedes or follows it. We offer a processing account of this effect of order on interpretation, which relies on the assumption that spatial coherence in film is established by INCREMENTAL VIEWPOINT GROUNDING.

## 1. Two Conventions of Film Interpretation

The basic art of film is to render the visible and audible “surfaces” of an action. But film, to tell stories, must get beneath the surface too. It has to convey what each character is thinking and feeling; what motivates one, and what another is left in the dark about. We can agree that a well-made film accomplishes this with natural ease, but how is it done? One way is general to the dramatic arts. We perceive a character’s actions, including their verbal utterances, and infer from each a goal and an obstacle set in its way (which together prompt the action).<sup>1</sup> Another method is characteristically filmic. It involves the eyes, whose language we are biologically disposed to read and attend,<sup>2</sup> and which film can isolate by means of the close-up.

Seeing is knowing; and film plots a path to what a character knows by representing what they do and do not see. Did she look away as he made that desperate face? Then she has no inkling, yet, of his money troubles. Was there someone looking down from the stairhead all that time? Then a member of the household witnessed a murder! A film can represent someone seeing something,

---

1. Ball 1983: 25–31.

2. Carroll 1993: 127.

without fuss, in a single shot that includes both the person and the object of their glance (with continuous camera movement from one to the other if necessary). But film has also evolved two distinct *interpretative conventions* for conveying that what is pictured in a shot (which we will call the OBJECT SHOT)<sup>3</sup> is being perceived by an *off-screen* character.



**Figure 1: glance shot and object shot.** A glance shot depicts a subject looking at something off screen. An accompanying object shot depicts the object of their glance.

On one of these, the POINT OF VIEW (POV) convention, the camera's view is understood to be that of a character (who is therefore not typically pictured, and indeed could not, without the aid of a mirror, be completely pictured, in the shot). Instances of this convention (signalled using "masks" of various sorts on the lens) date back to early silent film, and make up some of the earliest occasions of editing.<sup>4</sup>

Notwithstanding how natural it is to use the view of the camera for the view of a person, and how readily the viewer grasps this device, the POV interpretation is still a convention of filmmaking.<sup>5</sup> It is characteristic of conventions that alternative regularities could have been accepted in their place,<sup>6</sup> and the POV interpretation is but one among several interpretations that the naked object shot could attract, both in principle – because we could instead have regularized a convention according to which the camera coordinates matched the occipital "view" projecting backwards from the character's head – and in actuality. Indeed, any POV shot (sans mask) has an acceptable alternative interpretation as an *objective* angle, where the viewpoint is unattributed and motivated solely by what it reveals.

The other way to represent the act of seeing appertains it into two shots, and hence is a convention of editing. In one shot, the character's glance off screen is

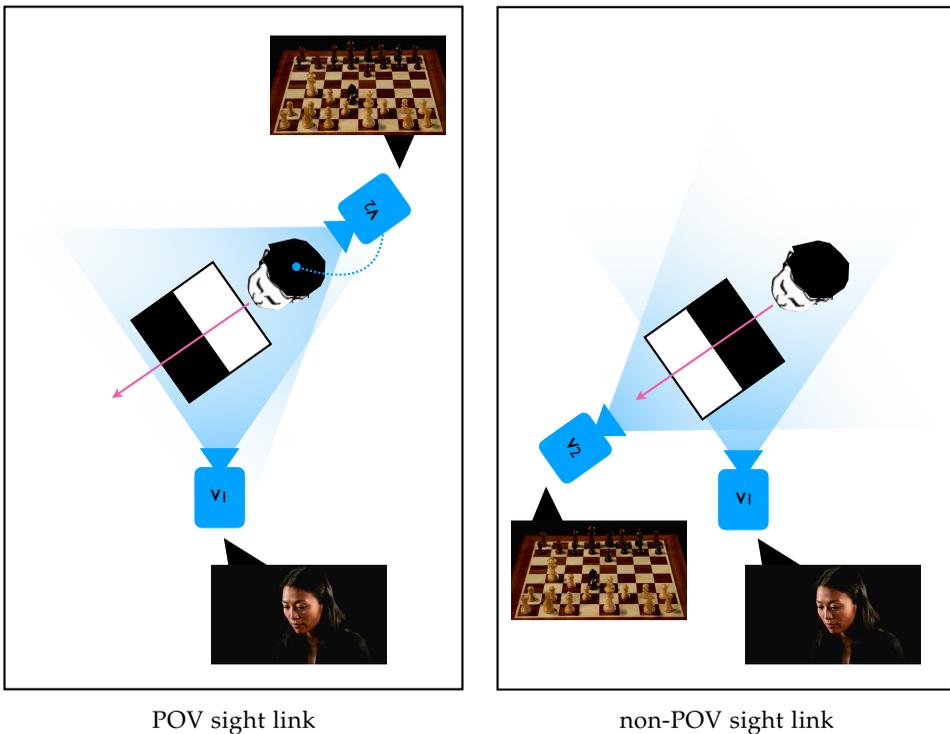
3. This is a compression of the terminology (POINT/OBJECT) used by Branigan in his classic paper on point of view (1975).

4. For instance, in G.A. Smith's *Grandma's Reading Glass* (1900). See Persson 2003: 50.

5. Albeit one that takes advantage of our normal tendencies. See Bordwell 2008: 57–82; Carroll 1993: 129.

6. Lewis 1969.

depicted (this is therefore called the *GLANCE* shot),<sup>7</sup> while in a distinct shot, the aforementioned object shot, the target of this glance is isolated by the frame (the glance itself is not visible in this shot). This sort of editing transition is sometimes called a *SIGHT LINK*.<sup>8</sup> It can be used in conjunction with a *POV* interpretation of the object shot, but the two conventions are independent and dissociable. A *POV* object shot may occur without an accompanying glance shot, and an object shot may be joined to a glance shot in a sight link without the angle of the object shot matching that of the glance.



**Figure 2: sight link with and without POV.** When the object shot is interpreted according to the *POV* convention, the camera's view is understood to coincide with that of a character. When the sequence as a whole is interpreted according to the sight link convention, the object shot reveals the target of the glance in the glance shot (but not necessarily from the angle of the glance).

In the next two sections, we attempt a precise description of both *POV* and sight link conventions. Under scrutiny, we find that these techniques for communicating a character's mind are not *intensional*, like their nearest counterparts

7. Branigan's (1975: 55) original terminology was 'point/glance'. The truncated versions are used, for instance, by Persson (2003).

8. E.g. Gunning 1991: 169.

in language, but geometrical. Indeed, POV and sight link are both VIEWPOINT CONSTRAINTS – conventional associations that underwrite the spatial coherence of an edited scene. Sight link itself turns out to be a special case of the familiar X-CONSTRAINT (Cumming et al. 2017).

### 1.1. POV

POV is an interpretative convention that applies to the individual shot. While the POV interpretation may be suggested by an accompanying glance shot (more on this later), it can occur without such a shot, and indeed without editing at all. In principle, a film could consist of one continuous POV shot.

POV in film is comparable with certain kinds of *clause embedding* in natural language. In English, a sentence like ‘The children were playing’ lays out an objective claim, namely that an activity of a certain type – *play* – with specific participants – *the children* – was in progress. But when the same or similar clause is *embedded* under an (intensional) verb, such as ‘saw’, it serves to characterize what someone perceived or thought:

- (1) a. *Linda saw that* the children were playing.
- b. *Linda saw* the children playing.
- c. The children were playing, *Linda noticed*.

The English construction in (1a) works as follows. The clause ‘the children were playing’ denotes a propositional content  $p$ , and (1a) as a whole expresses the claim that Linda confirmed by perception that  $p$  was the case (most likely she had a visual perception from whose content  $p$  could be inferred). In other words, the sentence has a layered structure, in which the content  $p$  is subordinated to the intensional operator contributed by ‘Linda saw’ (which we will leave unanalysed):

- (2)  $\text{SAW}_I(p)$

It is tempting to view POV in film along the same lines, as contributing an extra tier to the shot’s interpretation. The object shot, just like the embedded clause, expresses a particular content. A continuous piece of video footage (we set aside sound) – or SHOT – has a content that may be decoded from the recorded pattern of light incident on the camera lens. Roughly speaking, a shot represents the visible appearances, and the immediate physical (or fictional) causes of those appearances, that play out before the camera over a certain period of time. The camera may be (actually or apparently) moving during the shot, in which case the shot will represent the goings-on in a continuously shifting region of absolute space.

Call the content of the object shot  $q$ . On an objective (non-POV) interpretation,  $q$  is the totality of that shot’s meaning; it “says” merely that some physical events,

the causes of those appearances, transpired. On the POV interpretation, however, there is more to its content than this. It says, additionally, that someone saw those things transpire from that very vantage point. Let's give the POV subject a name, *S*. On the POV interpretation, then, the object shot "says" that *S* saw *q*, or, more particularly, it says that the content of *S*'s visual perception during the relevant interval is, or is characterized by, *q*. To say so is to analyse the POV interpretation as introducing the same sort of intensional operator as the English verb *saw*:

(3)  $\text{SAW}_S(q)$

While it is tempting to treat the POV convention this way – as though it were semantically equivalent to embedding under *saw* – we will argue for a different approach. On our account, the POV convention instead identifies the *position* and *orientation* of the camera with the position and orientation of *S*'s gaze.<sup>9</sup> On this approach, the POV interpretation is not intensional, in the sense of relating *S* (directly) to an *intension* or content. We believe, rather, that the viewer makes a very natural (but unconscious) inference from the position of the gaze to the perceptual content obtainable from that position.<sup>10</sup>

So why prefer this account? In the first place, we would like to distinguish POV from other conventions in film that clearly do involve intensional embedding. For instance, consider a flashback prompted by a character's recounting of past events, and understood as representing the content of that character's testimony. Note that we can tell that a flashback (sometimes) functions this way – analogously to indirect discourse – from the rare cases where the testimony, so expressed, turns out to be false, as revealed later on in the film.<sup>11</sup> Another example of intensional embedding in film is where a shot, or even a whole scene or longer sequence, represents the content of a dream or hallucination. In either of these cases, we have an interpretation where the content of the film segment, *q*, is placed in the scope of an embedder, such as '*S* said', '*S* dreamt', or '*S* hallucinated'.<sup>12</sup>

Note that while dreams and hallucinations are aspects of a character's subjective experience, the camera is not obliged to occupy the egocentric coordinates of the dreamer in a dream sequence. Indeed, it is not uncommon to have the dreamer appear on screen in the midst of their own dream.<sup>13</sup> It is necessary, then, to separate the (intensional) convention by which dreams and hallucinations are rendered from the one used to represent a character's POV.<sup>14</sup>

The key property that serves to distinguish POV from filmic intensional

9. This corresponds to Abusch and Rooth's extensional analysis (2017: 89).

10. Cf. Abusch and Rooth (2017: 89–90).

11. Hitchcock employed this strategy in *Stage Fright* (1950). See Wilson 2012: 156.

12. See Abusch and Rooth 2017 for a formal semantics of such embedding in visual narrative.

13. Cf. Wilson 2012: 152ff.

14. Note that the two are blended in Abusch and Rooth's intensional analysis (2017: 93).

embedding is its failure to “nest.” A dreamer may dream that she, or another character, is dreaming; and this dream-within-a-dream would be represented in film by a shot or scene nestled within the first-level dream sequence.<sup>15</sup> We could also have, no doubt, a flashback presenting the content of one character’s recounting occurring inside another such flashback, so that the top-level character is understood as narrating another’s narration. But though it is possible, of course, to see *that someone sees something else* (e.g., *that the children are playing*), this is not what is conveyed by successive applications of the POV convention in film.<sup>16</sup> Edward Branigan’s (1975: 63) acute description of one such dovetailing sequence in *Psycho* (1960) bears this out:

[W]e see Marion inside her car glance (shot A) at a policeman outside the car who then glances (shot B) at her car licence plate (shot C). . . . One characteristic of this structure is that while we have seen something from Marion’s viewpoint, we have also seen something that she cannot see: the licence plate.



Figure 3: POV doesn’t nest. Sequenced POV shots from *Psycho* (1960).

Marion does not see what the patrolman sees; she can’t, from her position inside the vehicle. Hence shot C does not receive the nested or *doubly embedded* interpretation: Marion sees that the patrolman sees that the licence plate number is ANL-709. Rather – and in keeping with our alternative geometrical gloss – the camera in shot C is understood merely to occupy the position of the patrolman’s eyes (and the camera in B to occupy the position of Marion’s eyes). We attribute a perceptual content to Marion that follows from her taking the camera’s perspective in shot B, rather than attributing the content of the embedded sequence consisting of shots B and C (namely the content that the patrolman sees that the licence plate number is ANL-709).<sup>17</sup>

15. This recursive expressive opportunity was adequately exploited by *Inception* (2010) and its many subsequent parodies.

16. While it is what the English environment ‘He saw that she saw that . . .’ means.

17. As a referee points out, failure to nest is consistent with an analysis of POV as an intentional operator, so long as the operator is confined to a single level of embedding (the operator is not allowed to embed itself; or, on a possible worlds analysis, its own world

Let's begin now to elaborate our alternative analysis of the POV convention. Henceforth, the **VIEWPOINT** of a shot is the apparent<sup>18</sup> position and angle from which the shot is taken in narrative space. The position is a point in the space (at some distance and direction from each of the objects in the scene), while the angle is specified by giving three (orthogonal) directions within the space. Standardly, we give the orientation of a camera by specifying which way is camera-left, which is camera-up, and which is camera-forward (the direction in which the camera is pointed).<sup>19</sup> If all we gave was the forward direction, or z-axis, we still wouldn't know how the camera was rotated about that axis – we wouldn't have determined which directions in space were up and left, from the camera's perspective. Once we have specified any two of the axes, however, the third may be derived, since it is orthogonal to the plane they both occupy.

A shot viewpoint is thus conveniently modelled as a *coordinate system*, a mathematical object consisting of a point (the *origin* of the coordinate system) and three orthogonal vectors of equal length (corresponding to a step of one unit's length in the left, up, and forward directions).

The viewpoint of a person (at a particular time) may be modeled in the same way: as a tuple of a point (halfway between the character's eyes) and three orthogonal vectors specifying the (forward) direction of their gaze and which ways are up and left from the perspective of that gaze.<sup>20</sup> A character's viewpoint is also the origin of their **EYELINE**. We treat the latter as a pair consisting of a point (the origin of the viewpoint) and a direction (matching the z-axis, or forward

---

argument cannot be bound by a higher POV operator). While clear examples of intensional structures in film – the dream and speech operators just mentioned – do nest, it is reasonable to claim that the perspectival nature of point of view (in contrast with the nonperspectival proposition-embedding verb 'saw') makes a nested interpretation unintelligible, justifying a restriction to one level of embedding. In reply, it could be said that our spatial account has the virtue of predicting the failure of POV to nest, rather than stipulating it (no matter how motivated the stipulation). Still we grant the point that this argument is not decisive.

18. The viewer's distance to an object depicted on screen is estimated using depth cues (e.g., projection size if the object is familiar), and those cues can be manipulated without moving the camera (by changing the focal length of the lens – see, e.g., Block 2001). As a result, we need to distinguish the apparent location of the view on the scene from the actual distance of the camera from the filmed actors and props. This notion of viewpoint as the apparent vantage point in the space also extends naturally to computer animation, where no camera is used.

19. These are the positive directions along the transverse (*x*) axis, the vertical (*y*) axis, and the longitudinal (*z*) axis, respectively.

20. The way films, including animated films, are produced means that only viewpoints that are themselves coordinate systems may be represented by the viewpoint of a shot. This means that what we have, with deliberate vagueness, described as "the perspective of the gaze" must always be represented by up and left vectors that are orthogonal to the eyeline (though they are free to rotate in this orthogonal plane). We do not speculate on how true-to-life this representational constraint is (as we remain vague about what feature of a person's outlook it is supposed to represent). Thanks to an anonymous reviewer for bringing this detail to our attention.

vector, of the viewpoint). We will say that an eyeline *intersects* an object just in case that object lies on the half-line, or ray, extending from the origin in the forward direction.

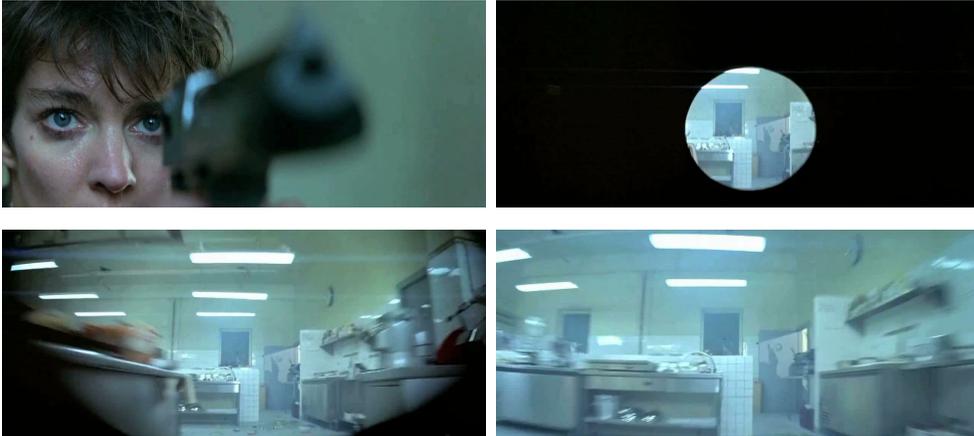
We can put the basic spatial meaning of the POV convention in a rough way as follows: the (possibly moving) viewpoint of the POV shot coincides with a character's (possibly moving) viewpoint. More carefully:

(POV) For a POV shot  $\sigma$ , a character  $S$ , and all times  $t$  represented by  $\sigma$ , if  $v$  is the viewpoint of  $\sigma$  at  $t$ , then  $S$ 's viewpoint is located at, and oriented as,  $v$  at  $t$ .

The POV convention can be seen as a special case of an interpretative convention that we might call OBJECT VIEW (OV for short). Object view extends POV by allowing the camera to represent the position and orientation of a non-sentient object that could not be said to project an eyeline. For instance, in one shot in *La Femme Nikita* (1990), camera movement evokes the trajectory of a bullet. Like POV, object view has been used since the early days of cinema. Thus, *A Kiss in the Tunnel* (1899) includes a moving shot from the front of a train as a tunnel looms. The more general convention may be stated simply by dropping the limitation to characters and their viewpoints:

(OV) For an OV shot  $\sigma$ , an object  $O$ , and all times  $t$  represented by  $\sigma$ , if  $v$  is the viewpoint of  $\sigma$  at  $t$ , then  $O$  is located at, and oriented as,  $v$  at  $t$ .

Note that OV portrays an object as having a particular *three-dimensional* orientation. If an object, such as a bullet, doesn't have an intrinsic "up" direction, then this would typically default to that of the scene space. One can also imagine a rotating OV bullet shot, where the up direction on the bullet is chosen arbitrarily, and the axial rotation of the camera portrays a corresponding spin on the bullet.



**Figure 4: object view (OV).** In this OV sequence from *La Femme Nikita* (1990), the viewpoint of the camera is understood to coincide with the changing location of a bullet.

Before moving on, we should pause to consider a common sort of counterexample to the account just given of POV. Sometimes the viewpoint of a POV shot appears to be placed, or else to move, *forward* of the position occupied by the POV subject. The successive POV frames below, from *Vertigo* (1958), illustrate the phenomenon:



**Figure 5: POV and framing.** In this sequence from *Vertigo* (1958), conventions for POV and framing combine to indicate a shifting zone of attention.

We diagnose this as interference from a related convention for achieving a refined sense of what a character sees. The convention employs the *framing* of the shot as a means of indicating the POV character's *zone of attention*. Currently, our model of viewpoint does not incorporate framing, so we must extend it first by adding two *field-of-view angles*  $\theta_h$  and  $\theta_w$  (representing how far from the forward vector, in degrees, the field of view extends, with height and width specified independently). The convention we have in mind adds that the field-of-view angles of the POV shot represent the limits of the subject's current visual attention.<sup>21</sup>

21. Once again, we have a convention that is (understandably) vague on how attention in

Since the size of the viewing screen is fixed, and whatever is framed by the camera gets projected to the whole screen, it follows that changing the size of the frame will affect the size to which an object projects on the screen. This can, in turn, shift the apparent viewpoint of the shot, since the size of a projected object is an important cue for judging one's distance from it. To convince yourself of this, imagine an alternative convention where the zone around some part of the screen is *masked* as a means of confining attention to that part, while all the objects pictured keep the same on-screen size. This approach, by contrast, would convey close attention to one part while maintaining the viewpoint's (apparent) position.

If we want to state the spatial convention of POV in a way that allows for forward shifts due to attentional framing, we can say that the viewpoint of the shot,  $\nu$  (at  $t$ ), represents the orientation of  $S$ 's viewpoint, and a position either at or (directly) forward of that viewpoint, depending on whether  $S$  has a narrowed attentional focus or not.<sup>22</sup>

### 1.2. Sight Link

Sight link represents the act of seeing in two parts, across two distinct shots. The glance shot depicts the glance (the moment at which the eyes of the perceiver fix on something), while the object shot shows the object, event, or situation that is perceived. The eyeline, or path of the glance, thus overlaps both shots. The glance shot presents its origin, and the object shot shows a segment displaced forward of the origin, far enough along to reveal the target of the glance.

A key observation is that sight link, in the standard form it takes in contemporary continuity-style filmmaking, incorporates a constraint on the *direction* of the eyeline relative to the viewpoints of the two shots. Roughly speaking, the eyeline is required to *enter* the object shot from the side of the screen (left or right) opposite that from which it *exits* the glance shot.<sup>23</sup> This gives the impression that the spaces disclosed by the two shots are adjacent cells, viewed from the same side.

A bit more carefully, the sight link convention may be stated as follows:

- (SL) If two shots  $\gamma, \omega$  are connected by sight link, then:
- a.  $\gamma$  depicts the glance at the origin of an eyeline  $\epsilon$
  - b.  $\omega$  depicts some object, event, or situation that  $\epsilon$  intersects
  - c.  $\epsilon$  has the same "x-direction" (general direction either leftward or

---

its sense relates to the reality of visual processing, including its distinction between attentional focus and periphery.

22. More carefully, if  $\nu$  is the coordinate system  $\langle \vec{x}_\nu, \vec{y}_\nu, \vec{z}_\nu, o_\nu \rangle$ , then  $\mu$  ( $S$ 's viewpoint) has the same orientation ( $\vec{x}_\mu = \vec{x}_\nu, \vec{y}_\mu = \vec{y}_\nu, \vec{z}_\mu = \vec{z}_\nu$ ) and a (possibly different) origin  $o_\mu$  such that  $\vec{o}_\mu \vec{o}_\nu = k \vec{z}_\nu$ , where  $k$  is a nonnegative scalar.

23. See, e.g., Bordwell and Thompson 2010: 236.

rightward) in both  $\gamma$  and  $\omega$ <sup>24</sup>

The third clause is particularly important, because the same abstract exit-and-entry pattern is followed in other contexts, where the action that connects the two shots is not an act of looking.<sup>25</sup> Thus, if a character physically exits a shot on the left, they are likewise expected to enter the next shot, continuing the same movement, from the right. They cannot re-enter *moving rightward* without suggesting that they have changed course in the meantime.

Just as we saw in the last section on POV, there is a more general spatial convention that includes sight link as a special case. This convention (potentially) governs any action that is spread out over two shots, requiring that the left-right aspect of that action's orientation be maintained in both. For now we call the convention ACTION LINK (we will see in a moment that it has gone by another name):

- (AL) If two shots  $\sigma, \tau$  are connected by action link, then:
- a.  $\sigma$  depicts one part of an action that projects an action line  $\alpha$
  - b.  $\tau$  depicts another part of the same action along  $\alpha$
  - c.  $\alpha$  has the same  $x$ -direction (general direction either leftward or rightward) in both  $\sigma$  and  $\tau$

This concludes our initial presentation of two conventions for representing seeing in film. On our account, they constrain the position of a character's eyeline in relation to the viewpoints of one or more shots. The content of the character's perception (and hence what they know) is then derived by inference from the geometry of the eyeline in narrative space. We have noted in passing that each convention can be seen as a specialized version of some more general standard for representing the orientation of an object or its path through space. In the next section, we will step back and situate the ways of "representing seeing" in a general framework for representing connected space in film.

24. What we mean by clause (c), in detail, is:

If  $\vec{z}_\epsilon$  is the forward direction of the eyeline,  $\vec{x}_\gamma$  the  $x$ -axis of the viewpoint of  $\gamma$  and  $\vec{x}_\omega$  the  $x$ -axis of the viewpoint of  $\omega$ , then  $(\vec{x}_\gamma \cdot \vec{z}_\epsilon)(\vec{x}_\omega \cdot \vec{z}_\epsilon) \geq 0$ .

Observe that  $(\vec{x}_\gamma \cdot \vec{z}_\epsilon)$  is the  $x$ -coordinate of  $\vec{z}_\epsilon$  in the coordinate space of the viewpoint of  $\gamma$ . Its sign corresponds to the  $x$ -direction of the eyeline in  $\gamma$ : leftward if positive, rightward if negative, and neither if zero. The condition above is, therefore, satisfied in the following three cases: (i) the eyeline is leftward in both shots, (ii) the eyeline is rightward in both shots, or (iii) the eyeline is neither to the left nor to the right in at least one shot. The third case is discussed in film production manuals (Mascelli 1965: 92; indeed, it makes it into Bordwell & Thompson 2010: 246), where it is asserted that a shot with a camera "on the line" may be cut together with a shot on either side of the line, and hence may intervene to smoothly conduct the camera from one side of the action to the other. Note that the third case is also what makes the sight link convention consistent with POV (since a POV shot is on the eyeline).

25. Again, see Bordwell and Thompson 2010: 236–238.

## 2. Viewpoint constraints

The convention introduced earlier as *action link* is maintained, on the filmmaker's side, by a production standard known as the 180° RULE. This rule, included in all manuals of film production,<sup>26</sup> requires the camera to stay within the half circle of rotational space on one side of an extended action, preventing edits that would reverse the *x*-direction of the action line. There is, of course, no corresponding album of *interpretation* rules from which the film viewer gains instruction. Still, production manuals assume that violating the 180° rule has the potential to confuse and disorient the viewing audience.<sup>27</sup> Presumably we have been passively instructed by all the film and TV we have watched, and now expect (at some level not readily accessed) continued conformity to the rule. Filmmakers conform because we expect them to conform, and we expect them to conform because they have done so in the past. Indeed, this consistency in production and expectation yields benefits to both sides, since filmmaker and viewer have a common interest in the communication of spatial information, and if the viewer expects the filmmaker to follow the 180° rule, then this narrows down the possible overall spatial layouts of the action, providing extra information that may be missing, or uncertain, from the shots taken on their own.

Regularities that are maintained out of common interest and the mutual expectation that the different parties will continue to do their parts are conventions, in the general sense first adumbrated by David Hume and subsequently analysed by David Lewis (1969). Moreover, conventions for representing content, including spatial content, are semantic rules. Cumming et al. (2017) argue that film has such rules, and discuss particularly a class of rules termed VIEWPOINT CONSTRAINTS. A viewpoint constraint is a rule that restricts the position and angle of the viewpoint of a shot to which it is applied – similar to the way a lexical semantic rule restricts the denotation of an utterance to which it is applied<sup>28</sup> – typically by limiting the ways it can differ from another viewpoint. All of the conventions we have proposed so far for representing the act of seeing, and the generalizations thereof, are viewpoint constraints in this sense. POV restricts the position of the viewpoint of the shot to which it applies by requiring it to coincide with (or else lie directly forward of) another viewpoint: that of a character. Sight link restricts the angle of the object shot by requiring the eyeline to project the same *x*-direction in that shot as it does in the glance shot (and applies a corresponding restriction on the angle of the glance shot in relation to that of the object shot).

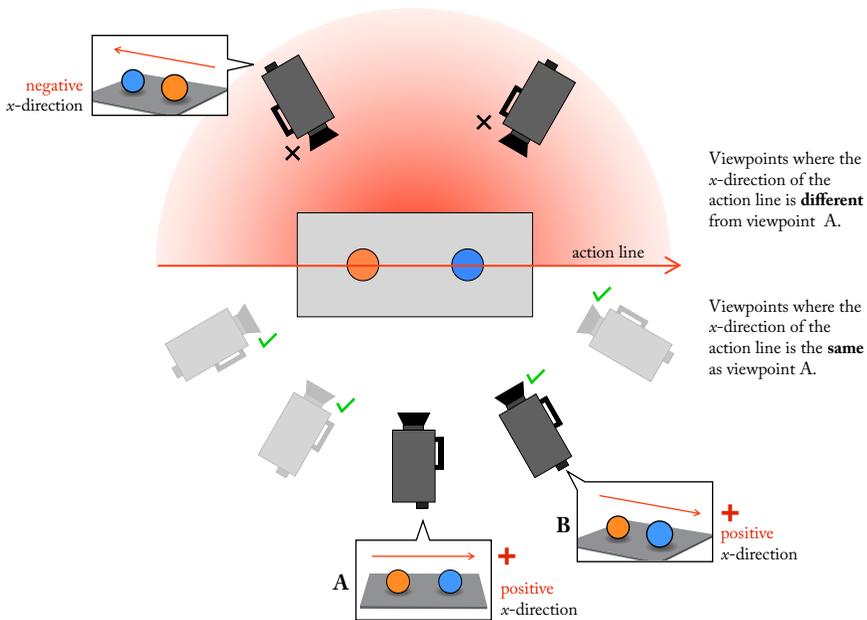
26. We have read our fair share. See, e.g., Mascelli 1965: 87ff; Miller 1999: 135ff.

27. See Kraft et al. 1991 for evidence that violating the 180° rule has a detrimental effect on memory of the scene.

28. Note that different lexical rules, constraining its denotation in different ways, could be applied to an utterance of the phonological form /bæt/. By the same token, the position from which a given shot is taken may be constrained by any one of a variety of viewpoint constraints.

Indeed, the viewpoint constraint behind the sight link convention was already discussed in Cumming et al. 2017. What we called “action link” a moment ago is in fact that paper’s central example of a viewpoint constraint – the X-CONSTRAINT (the title we will use henceforth).

- (X) If two shots  $\sigma, \tau$  are connected by the X-constraint, then:
- $\sigma$  depicts one part of an action that projects an action line  $\alpha$
  - $\tau$  depicts another part of the same action along  $\alpha$
  - $\alpha$  has the same  $x$ -direction (general direction either leftward or rightward) in both  $\sigma$  and  $\tau$



**Figure 6:** X-con. \*\*a problem is that the positive direction in the diag is rightward, but we say leftward in the text\*\*

### 2.1. Eyeline match and the R-constraint

One way of demonstrating that sight link is simply the X-constraint in the special case where the action line is a character’s eyeline would be to show that other (general) viewpoint constraints, when applied to an eyeline, yield distinct ways of representing seeing.

Let us consider, then, the phenomenon of EYELINE MATCH. Here is what one production manual has to say about it in the (standard) context of the “shot/reverse-shot” editing structure:

*Reverse shots* (also called *singles*) are closer shots of subjects in the scene. . . We can use the reverse of only one character in the scene or we can alternate between the reverse shots of both characters. Alternating between the two reverse shot angles is called *shot/reverse-shot* technique. . . An additional and important principle that applies to the camera placement for reverse shots is called the *eyeline match*. More exact than the general looking direction established by the  $180^\circ$  line (i.e., left to right or right to left), eyeline matching means being precise with camera placement and the focus of a character's gaze so that you accurately follow that character's sight lines from shot to shot, especially in an interaction. (Hurbis-Cherrier 2013: 75–76)

What is the stronger constraint imposed by eyeline match? This particular text doesn't tell us, though it is emphatic about the precision involved:

If you intend for there to be eye contact, the looking direction of a subject in a reverse shot must be focused *precisely* where the audience understands the other person (or object) to be. . . It is remarkable how just a little discrepancy can throw off the connection and prompt the audience to feel like the eye contact is askew. (Hurbis-Cherrier 2013: 76)

In other texts, including Katz 1991 and Arjon 1976, the stricter relationship is clarified under the general rubric of the “triangle system” of camera placement. In this system, two *symmetric* camera positions – the aforementioned reverse angles – form the base of the triangle. Though they are oriented towards opposite poles of the action line, the size of the angle each makes to the line is the same. Thus, a  $3/4$  angle on one character will be paired with a  $3/4$  reverse angle on the character facing them, a  $7/8$  angle with a  $7/8$  angle, a profile with a profile, and so on.

As Katz puts it, while “the exact angle of the shot, composition and shot size are infinitely variable within the triangle as long as the line of action is not violated” (1991: 142), “it is common practice to maintain the same [angular] distance from the camera for sight lines in alternating close-ups of two or more actors” (1991: 184).

Let's now articulate what amounts, or what would amount, to a further convention of showing seeing in film: eyeline match.

(EM) If two shots  $\gamma, \omega$  are connected by eyeline match, then:

- a.  $\gamma$  depicts the glance at the origin of an eyeline  $\epsilon$
- b.  $\omega$  depicts some object, event, or situation that  $\epsilon$  intersects
- c.  $\epsilon$  has the same  $x$ -direction in  $\gamma$  and  $\omega$
- d.  $\epsilon$ 's angle towards the camera in  $\gamma$  matches its angle away from the camera in  $\omega$ .<sup>29</sup>

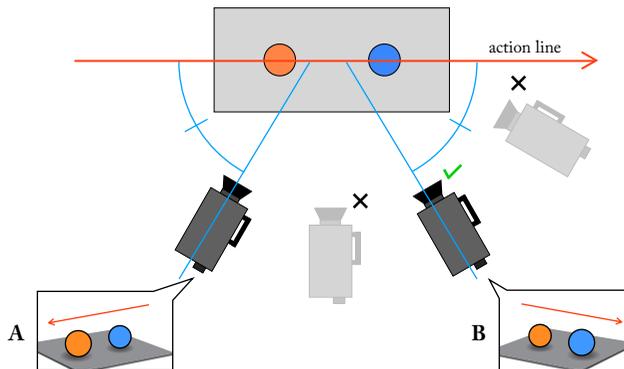
---

29. In short,  $\vec{z}_\epsilon$ , the forward vector of the eyeline, has  $x$ -coordinates that match in each shot

It will come as no surprise, by now, that the same mirroring of camera setups is employed in cases where the axis of action is not an eyeline. It is a further rule of filmmaking that a “soft” exit – when a character exits the frame at an angle that takes them close to the lens – should be matched by a similarly soft entrance from the other side of the lens.<sup>30</sup> Meanwhile, a “hard” exit – where the character exits at more of a right angle to the lens – should be followed by a hard entrance from the opposite screen-edge.

We dub the general constraint that applies when matching eyelines or entrances with exits the R-CONSTRAINT (for reversal/reflection).

- (R) If two shots  $\sigma, \tau$  are connected by the R-constraint, then:
- $\sigma$  depicts one part of an action that projects an action line  $\alpha$
  - $\tau$  depicts another part of the same action along  $\alpha$
  - $\alpha$  has the same  $x$ -direction in  $\sigma$  and  $\tau$ , and
  - $\alpha$ 's angle towards the camera in  $\sigma$  matches its angle away from the camera in  $\tau$ .



**Figure 7: R-Constraint.** The angle towards the camera in shot A matches the angle away from the camera in shot B. \*\*mistake in images: x switched instead of z (also action line needs to point left)\*\*

and  $z$ -coordinates that sum to zero. That is,  $(\vec{x}_\gamma \cdot \vec{z}_\epsilon) = (\vec{x}_\omega \cdot \vec{z}_\epsilon)$  and  $(\vec{z}_\gamma \cdot \vec{z}_\epsilon) = -(\vec{z}_\omega \cdot \vec{z}_\epsilon)$ . Note that while sight link is consistent with a POV interpretation of the object shot, eyeline match is inconsistent with it, except where the camera is on the eyeline in the glance shot (the glance could be into the lens of the camera, or else directly above or below it.)

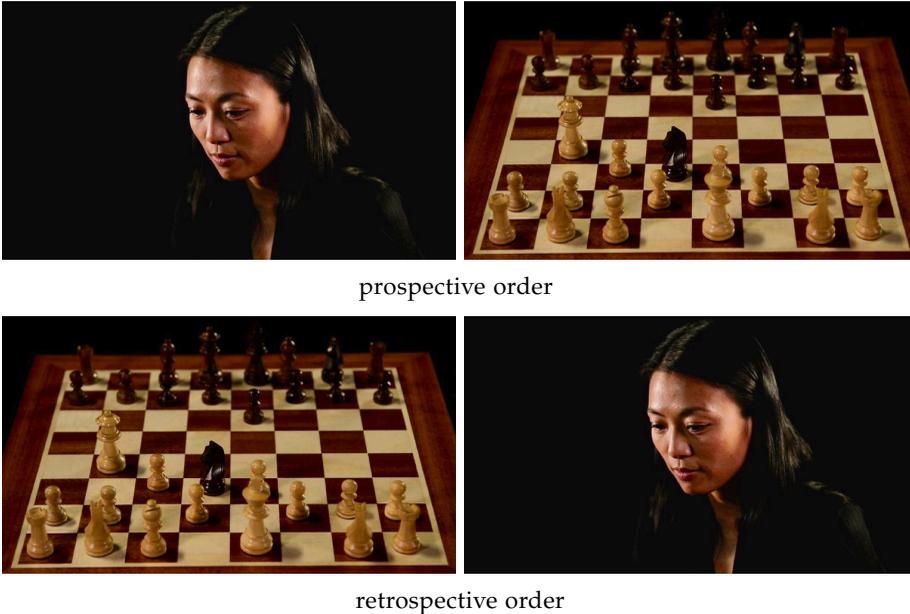
30. The rule is mentioned by Mascelli (1965: 100):

An exit made close to the side of the camera should be followed by a shot showing the subject entering the frame in a similar way. If the subject enters the far side of the frame in the next shot, the audience will be distracted; because the distance is too great to cover between straight cuts.

Our claim here is that editing represents seeing with the help of general rules, such as the X- and R-constraints, for geometrically binding shots of the same action. It is important for this claim that such rules make ineliminable reference to an action line – potentially an eyeline – projecting through both shots. This is so for the X-constraint, where the action line is the boundary that the viewpoint cannot rotate past, and also for the R-constraint, which requires the viewpoint to rotate to a position an equal angular distance from the action line. If this were not the case – and it is easy to think up geometric constraints that do not reference an action line – then it might be natural to treat the narrative connection between two shots separately from their spatial connection. On the one hand, we might say, the shots are connected by showing a glance and its target; while on the other, their viewpoints obey such-and-such a geometrical constraint (the camera rotates exactly  $90^\circ$ , for instance). In that case, we could write down the narrative conventions on one list – including those for representing seeing – and the geometric conventions on another. As it is, however, we have but *one* list of viewpoint constraints, which combine geometric and narrative connectivity. So, for example, we can derive the fact that the woman seen glancing in one shot perceives the object shown in the other from the claim that the sequence obeys the X-constraint, with the woman’s eyeline serving as the axis of action. We don’t require any further narrative information to arrive at this interpretation.

### 3. Order

Where there are different conventional ways of relating the spaces shown in successive shots, the question naturally arises: how does an audience decide which convention is in force on a particular occasion, in linking a specific pair of shots? There is no simple answer to this question; as with disambiguation in natural language interpretation, the factors involved are likely to be open ended. Still, there is one important empirical observation to be made about the attribution of POV in particular, which we have not seen discussed elsewhere, and that is, on the face of it, surprising. It is that the order of the glance and object shots in a sight link sequence affects the tendency to interpret the object shot as POV. We believe that readers should be able to see the difference for themselves, by comparing clips in which the shots take different orders at the following address online: <https://vimeo.com/236153570>.



**Figure 8: POV and shot order.** Glance shot before object shot (prospective order) encourages a POV interpretation. Object shot before glance shot (retrospective order) makes a non-POV sight link interpretation more natural.

We hope you agree that when the glance shot comes first – in what is known as the PROSPECTIVE order – it is very natural to feel you are looking through the eyes of the character in the second shot. When the glance shot comes second – in the RETROSPECTIVE<sup>31</sup> order – the POV interpretation is more elusive. One naturally feels that the woman is, rather, looking on from the far side of the board, behind the black pieces.

We first discovered the effect, quite by accident, in the editing room. We were trying to create an objective sight link sequence from a particular pair of shots, but found that a POV interpretation came through instead, until we switched the order of the shots so that the glance came second. Since then, we have run two rounds of experiments on Mechanical Turk, with stimuli produced in-house. In both, we found that changing the order from prospective to retrospective decreased the proportion of POV interpretations. For example, in the second experiment, in the baseline “short glance” condition, we found that the prospective order favoured the POV interpretation above chance, with a median probability of 0.74 and a tight credible interval (0.64, 0.83). In the retrospective order, by contrast, we found a median probability of 0.37 with credible interval (0.27, 0.47).<sup>32</sup>

31. The terms are Branigan’s (1984: 111).

32. The analysis was conducted using a fixed effects Bayesian logistic regression model with low information priors. Note that we used permutations of a single clip (played in reverse

For completeness, we note that the clear difference between the orders was significantly diminished in a second condition where the glance was prolonged from the editorial norm (about 2s) to an emphatically long 5s (this was accomplished by slowing down the original footage). We don't attempt to account for this effect below.

The analyses of POV and sight link in §1 do not incorporate sensitivity to shot order. They (correctly) allow either interpretation – POV or objective – to occur in either order,<sup>33</sup> and say nothing that would lead one to expect that placing the glance shot first in sequence should improve the POV interpretation for the object shot.

In the remaining sections, we discuss a proposal by Noël Carroll that may appear relevant to the effect just outlined, but which we show is unrelated. We then proceed to our own account of the order effect, which is based on the assumption that the narrative space of a film scene is built up *incrementally*.

### 3.1. Carroll on “Point of View”

In an important article on point-of-view editing, Noël Carroll (1993) connects the editing structure to an everyday perceptual behaviour: following the gazes of others to their objects. Gaze following is a standard way of finding out what others know, by seeing what they see (and that they are seeing it). As we have discussed, this also tends to be how film viewers come to know what characters in the film know – and even the cognitive mechanics turn out to be broadly similar.

Gaze following is a kind of triggered visual search. Seeing someone's eyes as they fasten on something acts as a DEICTIC CUE that involuntarily shifts one's own attention to the object of their gaze. The cue can be thought of as the perceptual equivalent of raising the question “what are they looking at?” (Hochberg & Brooks 1978).

In real-world vision, the viewer would then use their cued attention to either locate an object in the periphery of their vision or move their head to locate an object out of view. They would then perform a saccadic eye movement to the first object they found that aligned with the [observed] gaze. In film, the same projection of the gaze through visual space will occur but it will stop as soon as it reaches the screen edge. If the target

---

for the retrospective order) as the sole stimulus for each condition. Properly speaking, then, we have not demonstrated a general effect of order across a range of examples, but only experimentally confirmed the intuitions reported above for the clips you can watch online.

33. (POV) doesn't mention a glance shot, but only a character whose glance is represented (who, consistently with the convention, may never appear on screen). And while (SL) does mention both shots, it doesn't require them to occur in a particular order; it doesn't even require them to occur successively (sometimes, indeed, further shots are inserted between the two parts of a sight link).

of the gaze is found within the screen a saccadic eye movement will be initiated. . . . If no valid target exists the editor will have to provide one by cutting to the [object] shot. The object depicted in the [object] shot can either be located along the path of the actor's gaze, requiring a saccade to fixate. . . , or be collocated with the viewer's current point of fixation. . . . In the latter case no saccadic eye movement is required but attention will still be captured by the sudden onset of the expected object. (Smith 2006: 67)

Carroll takes the editor's exploitation of this common attentional pattern to be important for understanding film's mass appeal:

Stated baldly, point-of-view editing can function communicatively because it is a representational elaboration of a natural information-gathering behavior. That is, point-of-view editing, of the prospective variety at least, works because it relies on depicting biologically innate information-gathering procedures. This is why the device is so quickly assimilated and applied by masses of untutored spectators—or so I hypothesize. (1993: 130–131)

Carroll specifies “the prospective variety” of editing, by which he means a sequence where the glance is shown first, and its object second. It is only in the prospective order, of course, that the perceptual question is raised prior to the editor's provision of the answer. Carroll does not deny that the retrospective order is employed in film, but his explanation of the mass appeal of this kind of editing only really works for the prospective order, and so he has a theoretical motive for regarding this order as basic and the other as a derivative variant.

If the prospective order is basic and, moreover, conforms to the question-answer structure of gaze following, doesn't that explain why POV is more natural in that order? Actually, it is not at all clear that it does. Recall that POV is competing with an objective sight link interpretation, and so what we are looking for is a consideration that would make us prefer POV to objective sight link specifically. So if the prospective order turns out to be basic for both POV *and* non-POV versions of sight link, we don't yet have a reason to expect the POV interpretation to gain ground on its alternative when that order is followed.

But, if you think about it, the analogy with gaze following is more exact for the objective form of sight link. We do not, in real life, come to occupy the viewpoint of the other party when we track their gaze to its object. This is not to fault Carroll; indeed, what he *means* by “point-of-view editing” is, in our terms, sight link rather than POV.<sup>34</sup> His hypothesis, then, is that the prospective order is

---

34. This understanding of the phrase “point-of-view editing” is not uncommon in film theory. For instance, Bordwell and Thompson (2010: 244–245) specify “optical point of view” when they mean POV specifically.

basic for sight link *in general*, since only in that order is the editing convention a “representational elaboration” of natural gaze-following behaviour. Understood this way, Carroll’s point is of no help in explaining the effect of order on the preference for POV that we have observed.

### 3.2. *Analysis of the order effect*

How, then, to account for the effect of order on POV? One way would be to fill out the conventions further. For instance, we could *require* POV in the prospective order, and bar it in the retrospective (adding a meta-rule that guesses a shot is objective unless it immediately follows a depicted glance). While this predicts a bias in the right direction, it makes shot order fully predictive of POV, which is too absolute. There are plenty of examples in film where a POV object shot precedes the glance shot,<sup>35</sup> and untold numbers (including shot/reverse-shot pairs) where glance precedes object yet the object shot is not POV.

We prefer to leave the viewpoint conventions in their current (symmetrical) formulation, and pursue a processing account of the order effect instead. The account is based on the idea that the spatial coherence of a scene is established by INCREMENTAL VIEWPOINT GROUNDING at each cut. What this means is that each new shot (after the scene’s opening shot) has its initial viewpoint related, by some conventional viewpoint constraint (POV, X-constraint, etc.), to a viewpoint that is already on record (in the viewer’s memory) and salient. Generally speaking, this will be the last viewpoint of the previous shot (for the X-constraint or R-constraint) or else the viewpoint of a character whose glance appears shortly before the cut (for POV).

Let’s see how we might arrive at a processing account along these lines. When we watch a film, we conceive of the camera as positioned within, or else traversing, a continuous narrative space (which may or may not match the dimensions of some actual place). When a scene is elaborated across a *series* of shots, the norm is to locate the viewpoint of each shot (the imagined position and orientation of the observer) somewhere within that narrative space. If this doesn’t happen, the scene is spatially incoherent; at a certain point we lose track of where we are. Though sometimes such incoherence is tolerated, or even desired, normally it is avoided.

A logical way to present a spatially coherent scene is to have an initial wide shot, or MASTER, display the whole region in which the scene will play out. With the relative positions of people and landmarks established, the scene location of subsequent, closer shots will be clear from who or what is shown in them. This technique has been called ANALYTICAL editing, from the way it analyzes (or carves up) space. Spatial coherence may also be obtained without a master shot, through

---

35. The easiest to bring to mind will be those of the shark in *Jaws* (1975).

the technique of CONSTRUCTIVE editing, which builds up an overall narrative space by combining a series of closer views.

Both methods require the viewer to relate each shot to the space of some other shot, either the master or a closer view that may not overlap with it. But what sort of relationship needs to be established for a shot to feel integrated with the rest of the scene? It does not seem necessary to assign the viewpoint of a shot precise coordinates within the coordinate system of another shot. A constrained but loose relationship – as when we know that one car is following another, but remain unsure of the exact distance between them – is sufficient for a satisfying sense of spatial cohesion. But we can't let just any spatial relationship count, as that would amount to no constraint at all (cf. Knott & Dale 1994).

We propose something in between. To count as spatially coherent, each shot must be connected to the rest of the scene by a conventional viewpoint constraint (POV, X-constraint, R-constraint, etc.). Such constraints correspond to familiar and, to a degree, expected ways of relating one shot to another; and while the relationship is thereby constrained in predictable ways, it is not necessarily pinned down to exact coordinates (viewpoint constraints other than POV do not specify how far the camera *translates* between shots, for instance). When we do have a sense of a (fairly) precise spatial arrangement, as in the objective interpretation of the clips linked to earlier, this is due to world knowledge and commonsense reasoning elaborating on the less specific relationship provided by the conventional constraint. This proposal parallels accounts of rhetorical coherence developed for linguistic discourse (Hobbs et al. 1993).

While not required by anything that has been said so far, it is natural to think of interpretation as *incremental*, so that the normal procedure is to ground each shot in one coming *earlier in the sequence*, whose contents are already on the discourse record (or, from a processing standpoint, in memory) as the new shot begins. This is the key to predicting the order effect.

Recall that, in our clips, a POV interpretation of the object shot competes with a non-POV, or objective, interpretation. On the POV interpretation, the viewpoint of the object shot is spatially anchored to the viewpoint of the *character* depicted in the glance shot. While on the objective interpretation, spatial coherence is established by a constraint that relates the viewpoint of one shot to the viewpoint of another (relative to the scene landmark of the eyeline).<sup>36</sup> For simplicity, we

---

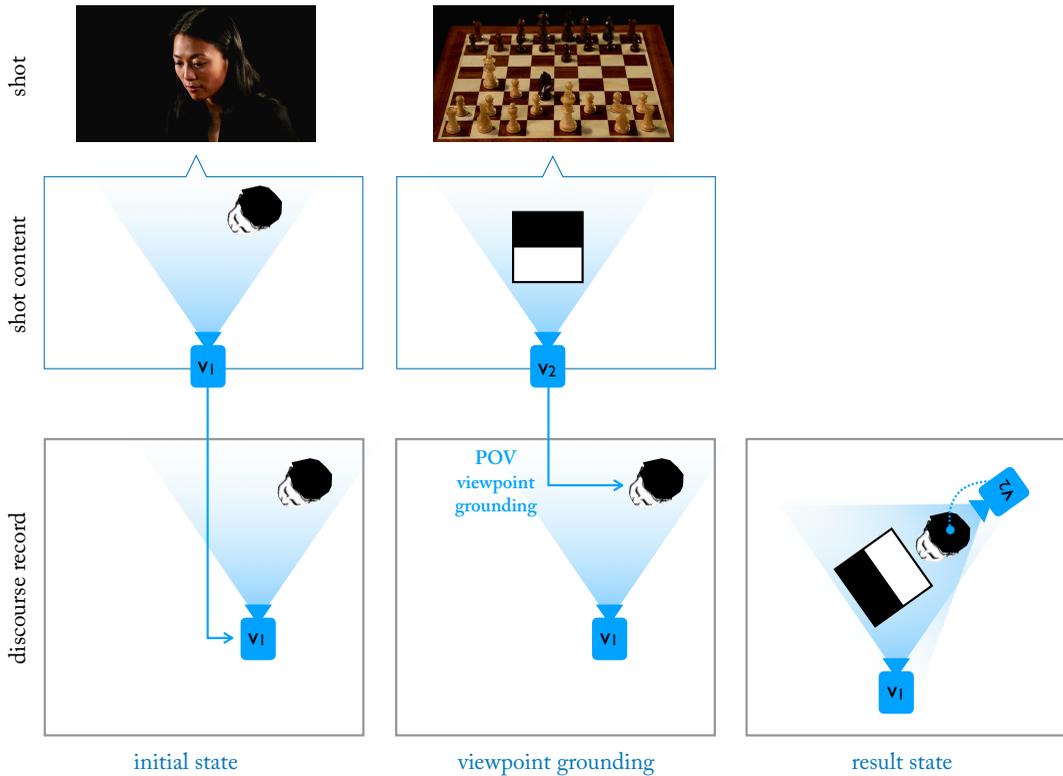
36. Incremental viewpoint grounding is the practice of assigning a viewpoint constraint and a viewpoint anchor (from the prior record of the film) to the initial viewpoint of each new shot. When the constraint is one that makes reference to an action line, that line will have to be located in both shots before coherence is established and the spatial construction of the scene is complete. But neither the action line nor its position in the anchoring shot will necessarily be on the record prior to the cut. Sometimes they will be, but other times the path of the action in the first shot will only be clear from watching the second (and connecting the two viewpoints with a geometric constraint).

assume this is the X-constraint, but the explanation to follow goes through for any similar constraint.<sup>37</sup>

Shots, without exception, have viewpoints. Hence the X-constraint (or similar) can always ground the viewpoint after the cut in the viewpoint that occurs right before it. The order of glance and object won't matter. With POV, however, things are different. Since it is the POV interpretation of the *object* shot we are concerned with, that shot must come second – following the cut – to have a viewpoint that requires grounding (and so *a fortiori* for its viewpoint to be grounded in the glance shot by the POV relation). In the normal course of things, we only expect POV in the prospective order.

---

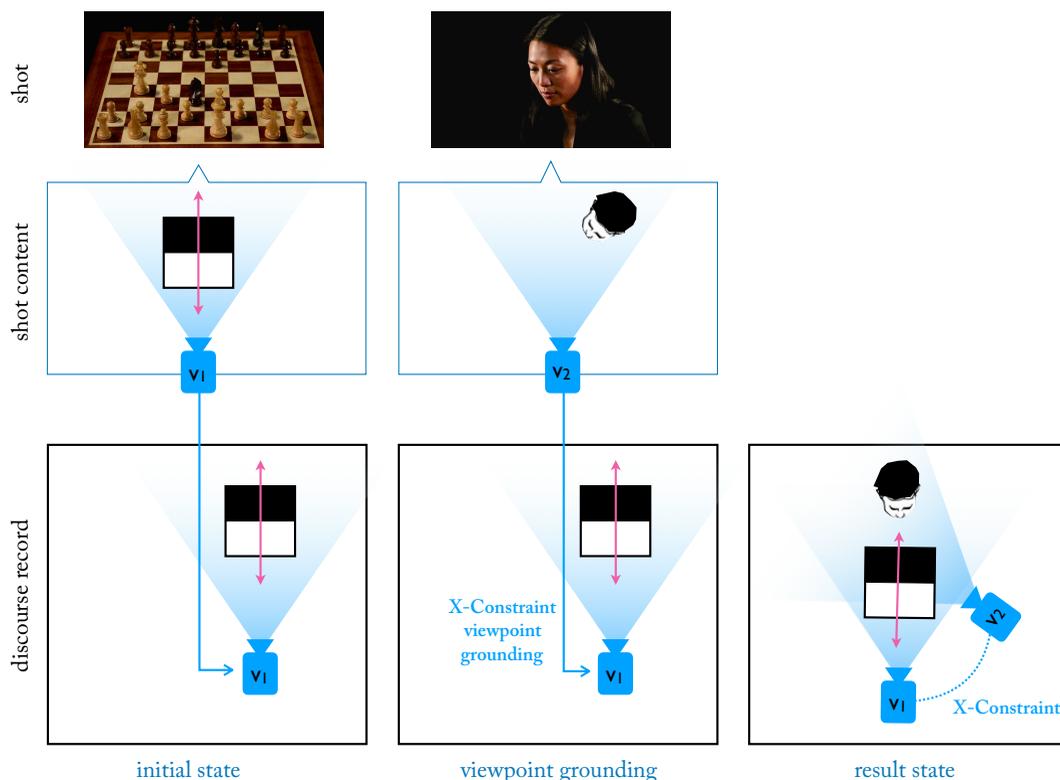
37. It can't be the R-constraint, as the camera angle on the woman is not mirrored by the angle on the board. The X-constraint could certainly apply on an objective arrangement (with the woman seated behind the black pieces), but it fails to *rule out* the alternative arrangement where she occupies the position of the camera (X-constraint is consistent with POV, as mentioned previously in fn. 24). For this we need another rule, one we might call the Z-constraint (not previously discussed, but the same as the X-constraint, only applying to the z-axis of the camera, and hence maintaining the general forward or backward direction of the action line between shots) or else Minimize Change (as discussed in Cumming et al. 2017).



**Figure 9: Incremental update for prospective order POV.** The viewpoint of the object shot (coming second) is spatially anchored in the viewpoint of the character depicted in the glance shot.

But that is only what's normal. The default tendency to ground each new viewpoint in the shot that has just ended can be overcome. For example, suppose we have a shot that we can tell is POV from internal evidence,<sup>38</sup> but whose location in the scene is unclear until a *later* shot, where the POV character appears on screen (like our retrospective order clip, except – we stipulate – the POV interpretation is already apparent while the object shot runs). This instance of spatial grounding is reminiscent of CATAPHORA in discourse, where a pronoun linearly *precedes* its antecedent. While the means of anchoring the shot (POV) is established at the outset, the actual anchor point in the scene (the POV subject's viewpoint) will not appear until a later shot.

38. This could be a mask, or some feature of the position or movement of the camera suggestive of an onlooker. Music and diegetic sound could also contribute.



**Figure 10: Incremental update for retrospective order non-POV.**

The viewpoint of the glance shot (coming second) is spatially anchored in the viewpoint of the object shot.

If this is a plausible example of cataphoric viewpoint grounding, it is also plausible that cataphora is not the usual method for establishing spatial coherence in film. It so happens that for certain pronominal relationships, in some languages, cataphora is the norm, and anaphoric resolution (where a pronoun initiates a search in memory for its antecedent) the more unusual case. But this can't be so for film. It would be strange indeed if the circuitous pattern of the hypothetical example just given was standard, while the straightforward approach of searching memory for an anchor point whenever a new viewpoint appears (i.e. after a cut) was comparatively rare.<sup>39</sup>

Cataphora is one way to derive the POV reading in the retrospective order. The other is *reanalysis*. The idea here is that while the first shot is playing the viewer defaults to an objective interpretation (does not apply POV), but then,

39. Note that if camera movement during a shot reverses the *x*-direction of the action line, then the next shot must match its *x*-direction by the end (rather than the one it started out with). This suggests that the viewpoint at the beginning of a shot is never anchored to that of a later shot by the X-constraint. Perhaps cataphora is only possible with POV.

after the cut comes and the glance appears, the viewer decides to revise their first-pass analysis and apply POV to the initial shot after all. Since the reanalysis occurs after the cut, the glancing character's position is salient and available as an attachment point. But reanalysis, which involves retrieving a previous interpretation from memory and making changes to it, creates an additional processing burden, and should be dispreferred as an interpretative option for that reason.

When all is said and done, the explanation we give of the order effect comes down to a preference for spatially integrating each shot *as it begins*, rather than delaying integration until *after it ends*. On our account, this entails settling on a coherence relation at the beginning of the shot, and initiating a search for a suitable anchor point. If no such point is found in memory, then resolution is put on hold until one appears (as in the cataphoric resolution of POV in the retrospective order) or the search is abandoned.

If you think about it, this assumption is a natural one to combine with our presentation of POV as a viewpoint constraint, and hence as one means of spatial integration. We do seem to assign POV at the beginning of a shot, as we can tell from the characteristic experience of viewing the action through the eyes of a character. If we only determined that a shot was POV once it was over, in the course of spatially integrating it with the shot to follow, the experience of POV would then be retrospective, rather than coincident with the action.

#### 4. Conclusion

We have engaged in a detailed study of POV and sight link, two conventions for representing seeing in film. In the course of this study, we have argued for several distinct points. First, POV should not be analysed as an intensional operator, along the lines of verbs of indirect discourse in language. Rather, the interpretation contributes a particular spatial constraint on the viewpoint of the shot to which it is applied, anchoring it in the viewpoint of a character (the origin of an eyeline). Sight link is a viewpoint constraint too, but one specific to editing. It anchors a shot in the viewpoint of another *shot*, with reference to an eyeline that serves as an orienting landmark in both. Indeed, sight link turns out to be a special case of the most common viewpoint constraint, the X-constraint. Other general viewpoint constraints, such as the R-constraint, produce more geometrically precise ways of connecting two shots by an eyeline.

Finally, we noted an effect of the order of the glance and object shots on POV. The POV interpretation is less accessible when the glance comes second. Rather than write further conditions into our definitions of POV and the X-constraint to encode this, we sketched out a way to derive the effect from the existing definitions. Our explanation turned on the fact that POV anchoring in

the retrospective (object-first) order requires the viewpoint of the first shot to be grounded in a location contributed by the second shot, whereas in the remaining three combinations of viewpoint constraint and order, the second shot could be grounded in the first. All we must assume, to predict the order effect, is that anchoring an earlier shot in a later one is less common and so less expected.

Of course, there is more to be said about the cues and conditions that favour the POV interpretation. For instance, editors have found that the timing of the cut relative to the fixing of the glance is important for continuity (see Smith 2006: 68–69), and our second experiment revealed an interaction between the length of the glance and shot order in the attribution of POV, so more than just the serial order of the shots must be considered in modelling this particular interpretative choice. Improved models would need to keep track of the salience of entities, both on-screen and in memory, as a function of time. Still, such models are recognizable (and realizable)<sup>40</sup> extensions of the basic idea of incremental update appealed to in our analysis here.

## Acknowledgements

Thanks to Ed Holsinger, Adrian Brasoveanu, and our film crew: Judy Phu (Director of Photography), Allyson Schwartz (Gaffer AC), Xue Lian Lei (Lead Actor). Thanks also to audiences at SCSMI, SPP, and Tübingen. This work was supported by a UCLA OVCR-COR transdisciplinary grant.

## References

- Abusch, Dorit and Mats Rooth (2017). The Formal Semantics of Free Perception in Pictorial Narratives. In Alexandre Cremers, Thom van Gessel, and Floris Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium* (85–95).
- Arjon, Daniel (1976). *Grammar of the Film Language*. Silman-James Press.
- Ball, David (1983). *Backwards and Forwards: A Technical Manual for Reading Plays*. Southern Illinois University Press.
- Block, Bruce (2001). *Visual Story: Creating the Visual Structure of Film, TV and Digital Media*. Focal Press.
- Bordwell, David (2008). *Poetics of Cinema*. Routledge.
- Bordwell, David and Kristin Thompson (2010). *Film Art: An Introduction, Ninth Edition*. McGraw-Hill.
- Branigan, Edward (1975). Formal Permutations of the Point-of-View Shot. *Screen*, 3(1), 54–64.
- Branigan, Edward (1984). *Point of View in the Cinema*. Mouton.

---

40. See Brasoveanu and Dotlačil (2020) for an appropriate modelling framework.

- Brasoveanu, Adrian and Jakub Dotlačil (2020). *Computational Cognitive Modeling and Linguistic Theory*. Springer.
- Carroll, Noël (1993). Toward a Theory of Point-of-View Editing: Communication, Emotion, and the Movies. *Poetics Today*, 13(1), 123–141.
- Cumming, Samuel, Gabriel Greenberg, and Rory Kelly (2017). Conventions of Viewpoint Coherence in Film. *Philosophers' Imprint*, 17(1), 1–29.
- Gunning, Tom (1991). *D. W. Griffith and the Origins of American Narrative Film: The Early Years at Biograph*. University of Illinois Press.
- Hobbs, Jerry, Mark Stickel, Paul Martin, and Douglas Edwards (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1–2), 69–142.
- Hochberg, J. and V. Brooks (1978). Film Cutting and Visual Momentum. In D. F. Fisher J. W. Senders and R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Functions* (293–317). Lawrence Erlbaum.
- Hurbis-Cherrier, Mick (2013). *Voice & Vision: A Creative Approach to Narrative Film & DV Production, Second Edition*. Focal Press.
- Katz, Steven D. (1991). *Film Directing: Shot by Shot*. Michael Wiese Productions.
- Knott, Alistair and Robert Dale (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1), 35–62.
- Kraft, Robert C., P. Cantor, and C. Gottdiener (1991). The Coherence of Visual Narratives. *Communication Research*, 18(5), 601–616.
- Lewis, David (1969). *Convention*. Harvard University Press.
- Mascelli, J. (1965). *The Five C's of Cinematography*. Cine/Graphic.
- Miller, Pat P. (1999). *Script Supervising and Film Continuity, Third Edition*. Focal Press.
- Persson, Per (2003). *Understanding Cinema: A Psychological Theory of Moving Imagery*. Cambridge University Press.
- Smith, Tim J. (2006). *An Attentional Theory of Continuity Editing* (Unpublished doctoral dissertation). University of Edinburgh
- Wilson, George M. (2012). *Seeing Fictions in Film*. Oxford University Press.