Spatial Coherence in Narrative Film

Gabriel Greenberg, June 29, 2022

Just as linguistic discourse is structured by relations of conceptual coherence, narrative film is governed by norms of spatial coherence. Spatially coherent scenes in film provide viewers with a vivid sense of space and orientation. Yet the nature and limits of spatial coherence remain largely unresolved. Drawing inspiration from the study of cognitive maps, I propose that film spaces take the form of abstract spatial graphs, with visual regions at their nodes. The resulting analysis helps explain what spatial coherence is, why some scenes have it and others don't, and how the 180° rule and other central principles of film making help promote it.

Section 1 describes the construction of film space and the role of the cognitive maps within a dynamic process of film interpretation. Section 2 contrasts spatially coherent and incoherent film sequences and asks after their essential difference. Section 3 develops a theory of spatial coherence within a graph-based account of film space. Finally, Section 4 examines conventional viewpoint constraints, including the 180° rule and POV editing, as strategies employed by filmmakers for achieving spatial coherence.

1 Film interpretation and the cognitive map

The interpretation of narrative film seems to work something like this. As a viewer watches a film, they construct a mental representation of an evolving situation or story. With each new shot, further information is incrementally added to the constructed situation, and this representational record grows in size and complexity until the end of the film. To interpret a film, or any other form of visual narrative, is not so much to recover a string of depicted moments, but to perform a sequence of updates to a central discourse record.[1]

In this way, film follows the mold of linguistic discourse, where interpretation is seen as a fundamentally constructive process.[2] One of the central questions for the theory of linguistic discourse is the ways that new contributions to a conversation can be **coherent** with the established record.[3] In taking up this question for visual narrative, we find that mainstream film and television exhibits many of the same forms of rhetorical coherence already familiar from linguistic discourse, including coherence with respect to narrative, causal, referential, and temporal relations.[4]

Acknowledgements. Special thanks to Mariela Aguilera, Joshua Armstrong, Sam Cumming, Catherine Hochman, and Rory Kelly for discussion and inspiration.

[1] See Abusch 2012; Wildfeuer 2014; Maier and Bimpikou 2019; Loschky et al. 2020; Cumming, Greenberg, Kaiser, et al. 2021.

[2] See e.g. Stalnaker 1978; Lewis 1979; Kamp 1981; Heim 1983; Gernsbacher 1997.

[3] See Hobbs 1985; Asher and Lascarides 2003; Kehler 2002; Asher and Lascarides 2003.

[4] See McCloud 1993; J. A. Bateman and Schmidt 2012; Wildfeuer 2014; Wildfeuer and J. Bateman 2016; Cohn 2018.

Yet in film, *spatial* coherence plays an especially central role.[5] Because individual shots convey such rich spatial information, there is opportunity for the relations of coherence between shots to leverage it. In particular, sequences of shots known as **scenes**, which express narratively unified chains of events, are bound by distinctive norms of spatial coherence. Scenes give narrative films their visceral sense of space and presence. This is why so many specific spatial conventions have arisen in film practice for shooting and editing scenes— the 180° rule, shot-reverse-shot, POV, jump cut, match-on-action, and eye-line match, to name only a few.[6] Meanwhile, marked discontinuities of space typically signal the end of one scene and the beginning of another. The aim of this essay is to identify the general, structural features that characterize coherent space in film scenes.

Here I am guided by a psychological conjecture: that the interpretation of film depends on two, interconnected capacities for spatial cognition. One is perception, which is centrally enlisted in the interpretation of individual shots. The other is the capacity to construct cognitive maps—evolving mental representations that track the spatial layout of the broader environment. The latter is used to glue together the contents of individual shots into a connected spatial lattice, the space of the film. The character of filmic spatial coherence, on this view, is defined by the kind of cognitive maps which viewers have the capacity to construct in response to watching a film.[7]

To be sure, in the context of film interpretation, we should not expect perception or cognitive maps to work exactly as they do in their ecologically native settings. They are adapted, constrained, and conventionalized for the demands of communication. They are subjected to genre and form-specific norms. As a result, the general psychological picture must be joined by a semantic theory, one which describes the specific, task-mediated form that spatial interpretation takes in the case mainstream, narrative film.

The core of the proposal developed here is that film space is built up in two stages. First there are the rich visual contents contributed by each shot, the counterparts of perception. Then there is a looser, graph-like network of linear spatial relations between these visual contents, what I call a **scene graph**, which is the counterpart of the cognitive map. Coherent scene graphs permit far more flexibility than traditional metric maps, but still impose substantive geometrical constraints. Scenes that do not fit into a coherent scene graph, I will argue, do not elicit a clear sense of space for the viewer; they may be marked as jarring or confusing, and may be difficult to remember.[8] As a result, much film work centers around providing viewers with the information necessary to

[5] See Kraft, Cantor, and Gottdiener 1991; Cumming, Greenberg, and Kelly 2017; Levin and Baker 2017.

[6] Burch 1981, pp. 10–11; Bordwell, Thompson, and Smith 2017, pp. 230–45.

[7] The map-based conjecture developed here is continuous with research on spatial encoding in film by Kraft (1987), Kraft, Cantor, and Gottdiener (1991), Levin and Wang (2009), and Levin and Baker (2017, pp. 9–11). While the idea that film interpretation emerges from spatial cognition is commonplace, most scholars have linked film interpretation with extended scene perception, rather than post-perceptual cognition of spatial layout. See e.g. Bordwell 1985, ch.7; Berliner and Cohen 2011; Cutting and Candan 2013; Tan 2018; Loschky et al. 2020.

[8] See Levin and Baker 2017, pp. 8–11 for a review of relevant research in psychology.

construct such a graph. Careful choices of camera angle, visual cues, and editing facilitate their construction at every turn.

The species of spatial coherence under investigation here is distinctive of shot-to-shot transitions *within* scenes. Transitions *between scenes*, while undoubtedly governed by their own norms of coherence, typically involve more open-ended narrative connections, and much looser spatio-temporal relations. Other forms of memory and conceptual representation beyond cognitive maps must be enlisted to interpret longer stretches of film narrative.[9] Likewise, there are a variety of non-scene sequences, such as montage sequences, which possess their own kind of coherence, to which the present account does not apply.

2 Spatial coherence

In mainstream narrative film, most scenes are **spatially coherent**: they give rise to a vivid and stable sense of space. Viewers know not just what is happening, but where. Spatial coherence is variously described in terms of the unity, continuity, or connectedness of the film space, or in terms of the viewer's ability to self-locate within that space: the viewer "stays oriented," "keeps their bearing," or "knows where they are."[10] I assume that spatial coherence is the default expectation for scenes;[11] dramatic ruptures in spatial coherence typically signal a change of scene.

The following sequence from *My Neighbor Totoro*, in which sisters Satsuki and Mei explore the attic of an old house, typifies robust spatial coherence. With every new shot, viewers can clearly and confidently locate the events depicted in relation to the previous action.[12]

[9] See e.g. Levin and Baker 2017, pp. 3–6, Tan 2018, pp. 9–14, Bordwell 1985, pp. 33–40.

[10] See e.g. Burch 1981, p. 10; Kraft, Cantor, and Gottdiener 1991; Berliner and Cohen 2011; Stork 2011.

[11] Berliner and Cohen 2011, p. 60.

[12] The one exception is shot 4, which depicts a swarm of "soot sprites" as they disappear into the woodwork. The shot momentarily elicits a kind of spatial confusion, since its relation to shot 3 is unclear. But this uncertainty is immediately resolved in shot 5, when we see that the sprites are the objects of Mei's backward gaze.

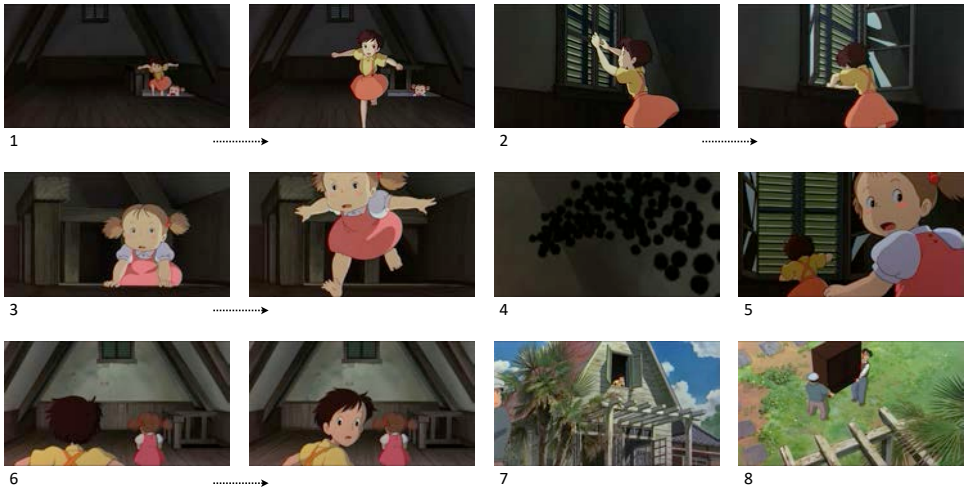

Figure 1: From *My Neighbor Totoro* (1988). Connecting arrows indicate when two stills are taken from the same shot.

When coherence breaks down within a scene, the sense of space collapses and disorientation takes hold. This may occur even when other aspects of the narrative remain coherent. Spatially incoherent scenes appear to leave viewers with vague or ill-formed representations of film space, and experimental evidence indicates that the resulting stories are difficult to store and to recall.[13] Such incoherence may be mistaken or intentional. When it is intentional, it is often used to enhance a sense of uncertainty, discomfort, or disruption in the narrative itself.

Consider this well-known sequence from *Breathless*. The scene is narratively coherent enough: it shows Michel Poiccard, the man in the hat, shooting a policeman. It is nonetheless marked by a palpable sense of spatial disorientation.

[13] See e.g. Frith and Robson 1975; Kraft, Cantor, and Gottdiener 1991; Levin and Wang 2009; Levin and Baker 2017.

Figure 2: From *Breathless* (1960).

Shots 1-3 comprise a spatially coherent subsequence: Poiccard looks up (shot 1), sees a policeman arrive (shot 2), then reaches into his car (shot 3). The disruption occurs in the transition from shot 3 to shot 4. In shot 4, we know we are looking at Poiccard, and learn soon enough that he is holding a gun. But we are left confused about the relationship between the space of shot 4 and the preceding space of shot 3. Did Poiccard turn around? Did he walk away from the car? How much time and space have been covered? The feeling of spatial incoherence elicited by these uncertainties reinforces the abruptness of Poiccard's fateful decision.

Cognoscenti will note that the sequence involves a violation of the 180° rule— in shots 1-3, Poiccard is oriented right-to-left on the screen, but in shot 4, left-to-right. (There may be another such reversal at 6/7.) This is true, and our upset expectations about screen direction partly explains why the cut is so jarring. Yet, as we will see, there are spatially coherent scenes which do preserve screen direction. So conformity or not with the 180° rule cannot be the whole story about the evident spatial incoherence at work here; I'll return to the proper explanation of this case in Section 3.

Isolated edits within scenes may be spatially incoherent without being marked as confusing, when the cues that would normally invite an assumption of coherence are voided. For example, cut away shots are often used to reveal an object within the setting of the main action, which is not itself spatially related to this action in any specific way. The central action might be two pool players in competition; a cutaway might reveal a clock on the wall. In such cases, no action line

purports to extend from the main action to the cutaway, and the cutaway itself doesn't create one. Though such shots do briefly subtract from the spatial coherence of the whole, the disruption is harmless, since, unlike the case from *Breathless*, the resulting spatial ambiguity doesn't concern the main action of the scene.

In recent years, however, spatial incoherence has become part of the editing style for mainstream action movies, where it often permeates entire scenes. The aesthetics of spatial coherence in action sequences has been the subject of heated debate among scholars and critics, and film enthusiasts on countless online forums.[14] Such scenes tend to elicit a sense of frenetic confusion. Some of this effect may be traced faster cuts and jerky camera work (Bordwell 2002). But as Stork (Stork 2011) argues film space itself is often incoherent in the new wave of action scenes.

In this scene from *Batman Begins*, for example, any concrete sense of space is gone.[15] We glean that there is a fight in an enclosed space, but little else. Whereas the scene from *Breathless* involved a single moment of mis-aligned action lines, there are simply no stable action lines to align here. This a deeper form of incoherence.

Figure 3: From *Batman Begins* (2005).

Describing such filmmaking as "chaos cinema," Stork captures the sense of spatial incoherence in cartographic terms:

> Even attentive spectators may have trouble finding their bearings in a film like this.

[14]Scholarly commentary includes Bordwell 2002, Stork 2011, and Shaviro 2016. Informal discussions on the web abound, with titles like: "Top 10 Movies with Incomprehensible Action Scenes"; "Why So Many Modern Movie Fight Scenes Suck"; "Why are movie fight scenes so bad?"; "How One Movie Trilogy Ruined Action Films Forever".

[15]The examples is taken from"Top 10 Movies with Incomprehensible Action Scenes" `https://www.youtube.com/watch?v=2wvYX-mlqKY`

> Trying to orient yourself in a work of chaos cinema is like trying to find your way out of a maze, only to discover that your map has been replaced by a reproduction of a Jackson Pollock painting. (Stork 2011)

So what distinguishes spatially coherent and incoherent scenes? In attempting to characterize the spatial character of film scenes, theorists tend to fall back on ideas of unity and continuity. Scholars have variously defined a scene as "a segment in a narrative film that takes place in one time and space" (Bordwell and Thompson 2010, p. 515); "a section of a motion picture which is unified as to time and place" (Katz and Nolan 2012); or "taking place in more or less continuous time and space" (Beatty 2004).

Taken as accounts of spatial coherence for scenes, these definitions are suggestive, but limited. No action takes place in just *one* location, narrowly construed; even shot-reverse-shot and POV editing build up space through the depiction of at least two, often discontinuous, spaces, like the sequence from *Vertigo* below: the character of Scottie Ferguson looks on (shot 1), while, across the pier, Madeleine Elster stares in the San Francisco Bay (shots 2 and 3). And of course, if locations are considered more broadly, they no longer capture a distinctive feature of scenes, since different scenes can take place in the same broad location.

Figure 4: From *Vertigo* (1958).

Taking inspiration from the idea that scenes are represented on a cognitive map, one might instead think that scene space is encoded by a kind of map, its constituent shots unified by location within a common coordinate frame. Cognitive maps are classically thought of as **metric maps**, locating every object they represent at a particular distance and direction from every other, just as they would be in a physical space. And there is considerable evidence that, at least for certain kinds of navigational tasks, humans maintain and update cognitive metric maps.[16] The idea that spatial coherence is defined by location on a metric map seems to be what Berliner and Cohen (2011) have in mind when they say that "spatial coherence indicates physical connectedness" (56): even though two shots may reveal strictly disjoint spaces, each overlaps with a single continuous spatial model (54-56).

Many kinds of scenes exhibit this kind of metric spatial coherence. To a first approximation,

[16]For reviews, see Gallistel 1990, Peer et al. 2021, and Epstein et al. 2017.

metric information is what is expressed by a single shot. So the use of a wide shot of an unfolding situation can establish metric relations between all the key objects in the scene. As a consequence, **analytical editing**— an editing technique in which close-ups always follow a master shot of the entire scene— tends to convey metric scene information, consistent with the metric map hypothesis.

But this is too strong a requirement for film in general: not all spatial coherence can be assimilated to metric connectedness, because there are coherent scenes where distance is indeterminate. We are typically presented with the beginning of an action in one shot, and the end of an action in another, but have no exact sense of the distance traveled between them.

In this scene from *Sesame Street*, we first see Abby throw a stick for Elmo's dog Tango. In the second shot, we see Tango running to catch it. Only, we don't have a determinate sense how far Tango has run when he catches the stick. We might have an approximate sense of distance, given world knowledge about how far children can throw sticks. But were we to plot the scene out in a physical model, the best we could do would be to indicate a range of possible locations for the second shot. Thus the scene space is not metric: there is no one coordinate frame or metric map relative to which all of the scene elements are assigned determinate locations. For all that, it is still perfectly coherent.

Figure 5: From *Sesame Street* (2021, S52E6).

This phenomenon is widespread. In **constructive editing**, viewers are presented only with closer shots of individual objects, and never provided with a master shot of the entire space. Thus, in many cases of constructive editing, the distance between shots is left open-ended. Consider again the POV sequence from *Vertigo*. The first wide angle POV gives us a sense that Ferguson is far enough away from Elster to take in the vista of the bridge. But the following close up reminds us that framing in POV isn't always a mark of distance. In the end, we know what Ferguson sees, but not how far away he is when he sees it.

As these cases show, a scene may be perfectly coherent while being indeterminate about distance. Thus it cannot be that the events making up a coherent scene are all located in a common metric frame of reference. And in general, some indefiniteness and uncertainty is compatible with spatial coherence. At the same time, as we've seen, extreme forms of spatial uncertainty do undermine coherence. Such considerations suggest that coherent scene space might be captured by a

relatively sparse spatial encoding, which nevertheless preserves a degree of geometrical structure.

These conclusions are anticipated by studies of human spatial memory, which show that, for certain kinds of tasks, the representation of the geography is far less rich than a metric map. To explain these findings, cognitive scientists have proposed that, in addition to metric maps, agents also posses a more flexible form of geographic representation, known as a **cognitive graph**.[17] One may imagine a variety of kinds of cognitive graph, but all involve nodes— typically locations or landmarks— connected by links— typically paths or transformations. Thus a cognitive graph might allow you to follow a sequence of turns to get to your destination, even if you could not compute the overall spatial relationship "as the crow flies" between your start and end points. Cognitive graphs emphasize the spatial connections between salient points rather than a common spatial coordinate frame. The introduction of cognitive graphs, I'll argue, gives us a way to understand the intuitions of unity and continuity in film while securing greater flexibility than permitted by metric maps.

In what follows I'll propose that a specific kind of cognitive graph models coherent space in film scenes. I take inspiration from Meilinger's (2008) idea that we think of the nodes of cognitive graphs as complete visual spaces, the result of discreet episodes of perception, and the links between them as rotational and translational transformations for moving from one visual space to another. Likewise, film space may be thought of as a kind of graph, with the visual spaces of individual shots as the nodes, and linear spatial relations as links between them. I develop this idea in the next section.

3 The structure of film space

This section first develops a general framework for understanding film spaces as graph-like structures can connect the contents of individual shots. I then outline an account of *coherent* film space, framed as a constraint on the connections between shots contents in a scene graph. I finally consider some less common alternatives to strict spatial coherence within the scene graph framework.

3.1 Scene graphs

Let us say that a scene (understood as a sequence of shots) **encodes** certain spatial relations between depicted objects when the combined information made available by the individual shot contents, world knowledge, and film conventions jointly entail those spatial relations. I will show how the spatial information that scenes encode can be realized in the formal structure of a scene graph.[18]

[17] See Peer et al. 2021 for a comprehensive review of recent literature.

[18] In carrying out this analysis, I will only consider spatial relations between objects as projected to the ground-plane of the film space. Such a 2D rendering of film space is of course a simplification, but it makes possible an elegant treatment of

The building blocks of scene graphs are **view spaces**: visual cones populated with objects and properties, whose positions are specified relative to a central viewpoint. For present purposes, I'll assume that view spaces are metric, governed by a local coordinate system whose origin and axis are fixed by the viewpoint.[19] Each shot contributes a distinct view space, with a distinct local coordinate system, over time, to the scene as a whole. To illustrate, let us revisit the first two shots of the coherent sequence from *My Neighbor Totoro* discussed earlier:

Figure 6: From *My Neighbor Totoro* (1988)

The diagram below shows the view spaces expressed by shots 1 and 2 taken separately. (Note that the orientation of the diagrams on the page carries no significance for the space represented.)

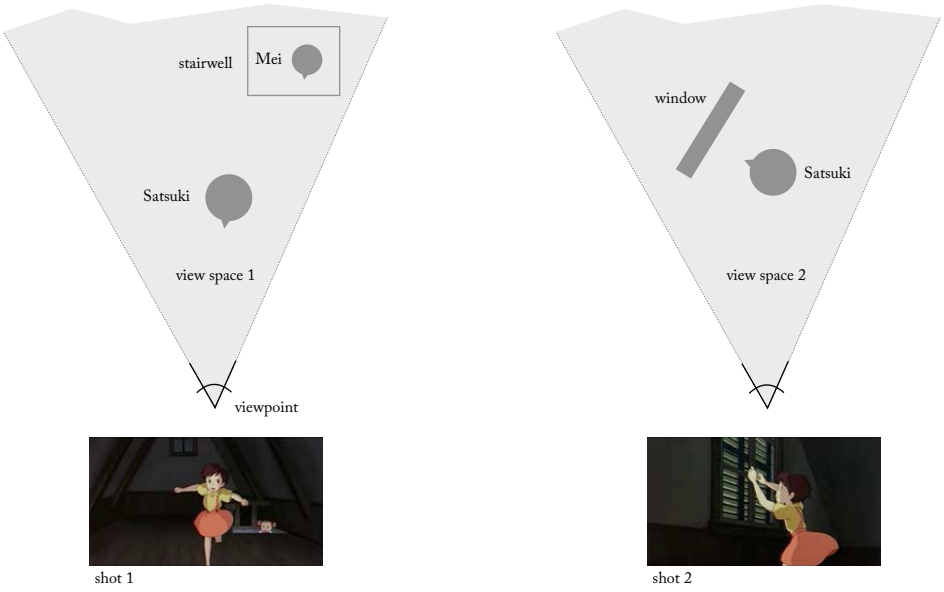

Figure 7: View spaces for two shots from the *Totoro* sequence.

At minimum, a **scene graph** is a collection of view spaces. Normally, these view spaces are

a wide range of cases. Here I follow the methodology of Cumming, Greenberg, and Kelly 2017 and Cumming, Greenberg, Kaiser, et al. 2021.

[19] This is a simplifying assumption; Greenberg (2020) argues that view spaces are not in fact metric, but have a geometry more like scene graphs themselves.

connected by a network of linear spatial relations that I call **leylines**, as in (8) below. Leylines are bounded at either end by an object, as located in a view space, and typically span more than one view space. Connection by leyline carries the content that there is a straight line which intersects the two objects in the world of the film, but this connection on its own makes no commitment about the relative distance or direction of the two objects to one another. (This is why leylines aren't simply line segments, which always have a length and orientation.)

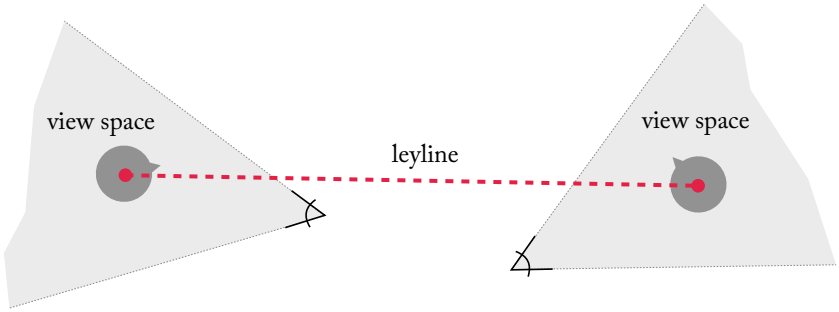

Figure 8: Two view spaces connected by a leyline.

Leylines may be supplemented with a variety of spatial relations, just as links may be labeled in a graph. When the distance between two objects is encoded by a scene, the resulting leyline may be labeled with a precise distance or with a range of distances. Likewise, the position and orientation of a leyline in a given view space may be left unresolved or fully determined. I'll use a red coordinate axis in the scene graph diagrams to indicate that the marked leyline has a determinate position and direction; see Fig. 9 for an example.

Note that leylines, in general, establish connections between *objects* in view spaces, not viewpoints or view spaces as a whole; just as objects are the focus of our attention in film interpretation, they are also the structural nodes in film space. After all, it is objects (including characters) which run, look, point, and throw sticks from one view space to another. So it is objects, not entire spaces, which we are normally in a position to put into linear relations.

Returning to the sequence from *Totoro*, we infer that Satsuki runs in a straight line from shot 1 to shot 2, following the direction of her gaze in shot 1. Thus we may construct a leyline connecting the two view spaces. It has a determinate direction at each anchor point, but the distance it covers is relatively indeterminate, hence unlabeled in the diagram below. This sequence is an instance of the general phenomena noted in the last section, where scene space, though structured in certain respects, falls short of full metric connectedness.

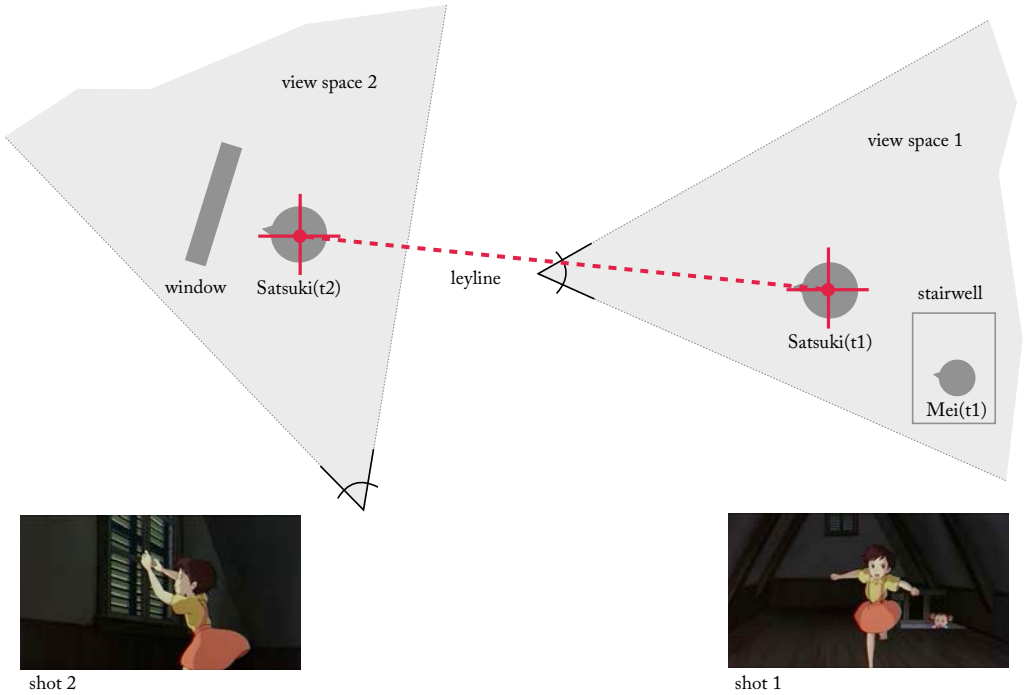

Figure 9: Scene graph for the *Totoro* sequence.

The construction of a scene graph is a dynamic process: as a scene unfolds, new view spaces are added to the established record. There is a default expectation that, whenever possible, a new view space will be connected by leylines to an existing view space. Such connectedness is a background requirement for spatial coherence. Here there is a strong preference for connecting with the most recently added view space, but the last two or three can be used. Sometimes a view space's position is not resolved, but kept in memory, until the next shot, which itself is connected to a previously established object.[20]

Linking the objects in a new view space by leylines to objects in a view space already on the record is in many ways like resolving the antecedent of an anaphoric expression. The interpretive engine assumes connectedness, and searches the recent record for a point of connection, on pain of incoherence. This picture of filmic "anaphora" offers an alternative to the account put forward by Cumming, Greenberg, Kaiser, et al. (2021, pp. 745–54). They hypothesize a default expectation that all *viewpoints* be resolved into definite relations with previously established viewpoints. The current proposal shifts the focus of dynamic update to the *objects* depicted. As I'll argue in Section 4, while there are indeed spatial conventions that connect viewpoints, such connections are driven

[20]Cumming, Greenberg, Kaiser, et al. (2021, pp. 745–54) discuss examples of this phenomena involving POV in detail. Shot 4 from the longer *Totoro* sequence in Fig. 1 seems to be a case in point.

by an underlying search for spatial relations between objects, and the latter is the ultimate arbiter of coherence.[21]

3.2 Absolute bearing

You ask for directions to the nearest gas station. "Go straight down this road until you reach the stop sign, then turn left and it'll be at the end of the block." These instructions are clear and coherent, but consider the kind of mental map that you construct in response. It marks the direction you must go down a road, the landmark at which you must turn, and the direction of the turn, but not how far the turn or the landmark is from your current position. This shows, at least, that though the mental map may require relatively determinate directions and rotations, it operates comfortably with highly indeterminate distances. Though metric space encodes the direction and distance of every object from every other, these two spatial features seem to come apart in cognition.[22]

The same seems to be true of film. Recall our examples of coherent but distance-indeterminate scenes from *Vertigo* (Fig. 4), *Sesame street* (Fig. 5), and *Totoro* (Fig. 6). In every case, although we don't know the distance between one view space and the next, we do know the *direction* of each view space relative to the other. Given the chance, we'd know how to walk from one view space to the next. I propose that this characteristic is also the central feature of coherent scene space. Roughly: for a film sequence to be spatially coherent, in each pair of shots, were you to enter into the space depicted in one shot and occupy the position of an object there, you could then point to an object in the other shot space, and visa versa. This is a relation I'll call **absolute bearing**: the scene encodes the bearing of each object relative to the other.

Within the scene graph framework, absolute bearing may be understood as a constraint on leylines: when each end of a leyline has a definite anchor point and direction, in two separate view spaces, then the two end points stand in a relation of absolute bearing. It will be useful to factorize the notion of direction into three components. We may think of a **direction** as a ray ℓ within a given view space v as defined by: (i) an **anchor point** p, the starting point of the ray; (ii) the **slope** of ℓ, defined as the set of vectors with which ℓ is parallel; (iii) the **sign** of ℓ, defined as the value, positive or negative, of subtracting the Y-coordinate of p from the Y-coordinate of any point along ℓ, given an arbitrary coordinate system for which ℓ is parallel to the Y-axis. Intuitively, the "sign" indicates whether a leyline goes in a negative or positive direction along its slope. These elements are illustrated in (10) below. When a scene encodes the anchor point, slope, and sign of a ray that intersects some object o, then it encodes the direction of o.

[21] That said, Cumming et al. use the viewpoint grounding assumption to provide a detailed explanation a specific set of data; it remains to be seen whether the present framework could successfully account for the same data.

[22] See Greenberg 2020 for analogous points about visual perception.

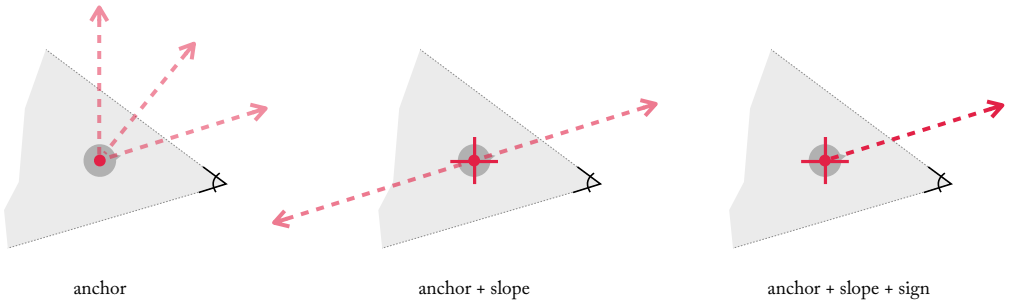

Figure 10: Components of direction: anchor, slope, and sign.

Following this schema, we may define bearing as follows:

(1) A scene S encodes the **bearing** of o_2 in view space v_2, relative to o_1 in v_1 iff S encodes that there is a half-line ℓ that intersects o_2 in v_2 such that:
   a. ℓ has o_1 as its anchor point in v_1;
   b. ℓ has a determinate slope in v_1;
   c. ℓ has a determinate sign in v_1.

We then define absolute bearing as the relation between two objects when the scene encodes the bearing of each object relative to the other:

(2) Objects o_1 in v_1 and o_2 in v_2 stand in a relation of **absolute bearing** in scene S iff
   a. S encodes the bearing of o_1 in v_1 relative to o_2 in v_2;
   b. S encodes the bearing of o_2 in v_2 relative to o_1 in v_1.

Finally, a scene graph as a whole establishes absolute bearing, or is **AB-coherent**, when every view space contains an object that bears a relation of absolute bearing to an object in another view space.

(3) A scene graph G expressed by a scene S is **AB-coherent** iff for every view space v_1 in G there is distinct view space v_2 in G such that there is an o_1 in v_1 and an o_2 in v_2 such that o_1 in v_1 and o_2 in v_2 stand in a relation of absolute bearing in S.

Absolute bearing is the residue of metric space when distances are removed or made optional. A metric space determines the distance and direction of every object relative to every other. Thus all pairs of objects stand stand in relations of absolute bearing. AB-coherent scene graphs abstract from metric space in two ways. First, they specify spatial relations between only a select set of the total objects they depict. And second, for those objects whose spatial relations are defined, they drop the necessary specification of distance. What AB-coherent scene graphs preserve from metric space is the fixation of mutual direction.

In the short sequence from *Totoro* (Fig. 6) it is clear that the resulting scene graph is AB-coherent, since we know that the two view spaces are connected by a straight line corresponding to the trajectory of Satsuki's movement, and we are given the position and orientation of both the start and end of that movement across the two shots. Readers are encouraged to revisit the original, longer sequence from *Totoro* (Fig. 1) to see how an extended scene may maintain AB-coherence throughout.

Insofar as individual view spaces encode metric information, they will almost always encode absolute bearing internally, between their constituent objects.[23] Likewise, the use of master shots in analytical editing typically encode metric information about the whole scene, thereby guaranteeing absolute bearing throughout. By contrast, constructive editing, as in the example from *Totoro* above, often results in spatial layouts that do establish absolute bearing, but otherwise fall short of metric completeness.

My proposal is that absolute bearing is the primary form of spatial coherence for film scenes. Scene graphs that do not establish absolute bearing correspond to spatially incoherent scenes. Coherence in this sense is something like a norm, rather than a conventional assumption, because viewers do not assume absolute bearing outright, but rather look for cues about how it is established, and make ancillary assumptions to derive it.

The equation of spatial coherence with AB-coherence sets both and upper and lower bound on spatial coherence. One one hand, it explains how scenes may be coherent even when they don't express complete metric structure, as we've seen with the examples from *Sesame Street*, *Vertigo*, and *Totoro*. AB-coherence is more lenient than metric structure.

On the other hand, film spaces which fall below the norm of absolute bearing feel incoherent and disjointed. To illustrate, let's return to the key sequence from *Breathless*, reproduced with new shot numbers below:

Figure 11: From *Breathless* (1960).

What's obvious first of all is that that the transition from shot 2 to shot 3 doesn't match our expectations: (a) we expect Poiccard to face the policeman, and (b) we expect screen direction to be preserved. A traditional analysis might diagnose the sense of incoherence for the sequence

[23] I say "almost", since shots which pass through complete dark patches, like those used in "single shot" films like Hitchock's *Rope* (1948), can interrupt spatial coherence.

as a whole as stemming directly from these violated expectations. The problem, as we'll see in Section 4, is that, under certain circumstances, such expectations can be foiled but spatial coherence maintained. The root of the incoherence, I believe, has to do with the interpretive fallout when the default assumptions fail.

When the default expectations (a) and (b) above aren't met, we are left with two salient interpretive possibilities: either Poiccard doesn't face the policeman and screen direction is preserved, or he does and it isn't. As the diagram in Fig. 12 shows, each interpretation corresponds to a distinct view space in a distinct relation to the previously established view spaces. Several shots later, the ambiguity is resolved, but by then the damage is done. Without enough information to establish absolute bearing on the fly, spatial coherence falters, and viewers are unable to construct a robust spatial representation of the unfolding events. We infer *what* happened, but we are left uncertain about *how*, spatially speaking, it came about.[24]

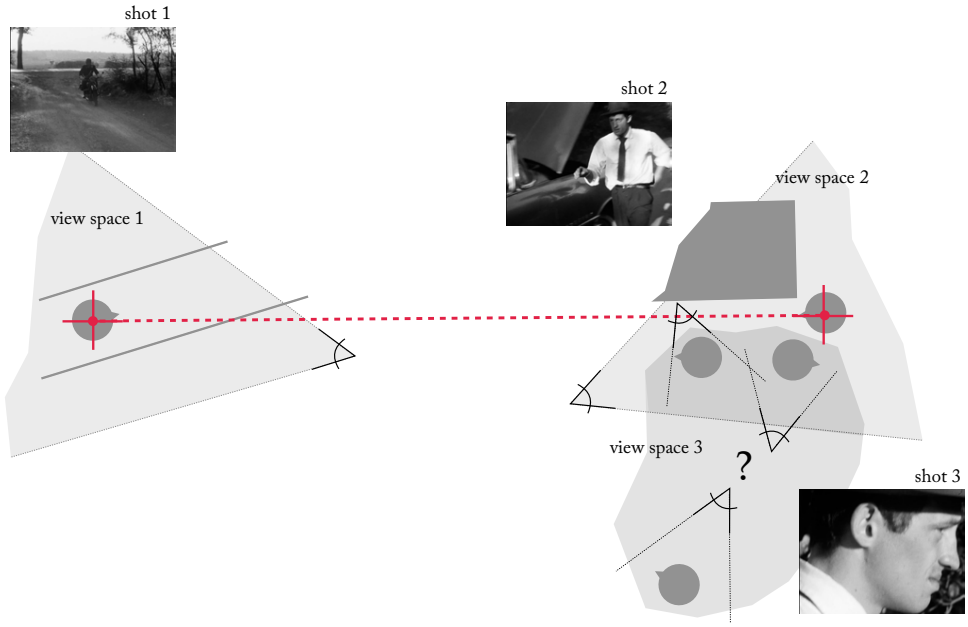

Figure 12: Unresolved scene graph for the *Breathless* sequence.

Meanwhile, in the scene from *Batman Begins* shown above, the spatial incoherence is even greater. There is no ambiguity about how to resolve the lines of action because there are no lines of action to resolve. The scene graph consists of a set of view spaces which, very approximately,

[24]The ambiguity is confounded because there is jump in time between shot 2 and shot 3: in shot 2, Poiccard is bending over to reach into his car, and in shot 3 he is standing up outside the car. (See Cumming, Greenberg, and Kelly 2021 for discussion of time jumps.) We don't know exactly how much time has elapsed or how much Poiccard has moved during that time, resulting in even greater interpretive uncertainty.

are organized into a loose radial formation. We never know the exact directional relation of one to another, as no leylines between the view spaces are ever established.

3.3 Varieties of coherent space

I've argued that absolute bearing is the primary form of spatial coherence for film scenes. But there are other spatial relations, less frequently used, that seem to support coherent spatial interpretation, albeit of a looser texture. Here I'll consider some of the options that are systematically related to absolute bearing.

A prominent editing pattern which we might call **radial editing** involves a shot of a central object intercut with shots of individuals in a group (often in a circle or line) looking at the object. We infer that each is looking at the same thing, but we are not given enough information to put together the spatial relations *between* the onlookers. Here's a typical example from *The Great Pottery Throwdown* , with the judges at the center, and the participants waiting anxiously for their judgement:

Figure 13: Radial editing in *The Great Pottery Throwdown* (2020, S3E3).

In radial editing, a series of leylines are introduced for which the anchor, slope, and sign of the line are fixed at one end, but only the anchor is fixed at the other. The former is contributed by any one of the glance shots, while the latter is contributed by the object shot. We may think of this as a case of **partial bearing**, where one object has the bearing of the other, but not visa versa.

Radial editing is facilitated by a variant on POV editing that Cumming, Greenberg, Kaiser, et al. (2021) call **sight link**: a sight link connects a person with a gaze to the object they are looking at, but *not* necessarily from their point of view. Without a master shot, repeated use of sight link to the same object results in a radial configuration like the one below.

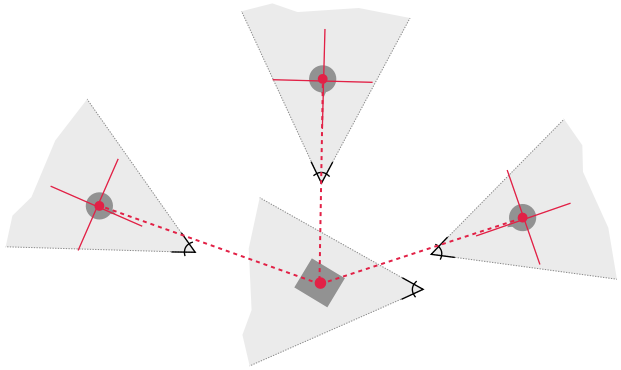

Figure 14: Scene graph for radial editing: a leyline from each peripheral view space is anchored on a central object at an indeterminate direction

A second kind of editing pattern which we might call **alignment editing** involves the depiction of disjoint lines of action, with different characters, all understood to be oriented in the same direction. This is a common technique for filming battles, like the scene from *Lord of the Rings* below. Here, a sequence of shots show warriors up and down the battle line, but all are interpreted to be moving in a common direction. Consistent screen direction (leftward or rightward) helps us keep track of which side is advancing in a given shot. But what is conveyed by the scene is something richer: the alignment of direction of motion in the depicted world.

Figure 15: Alignment editing in *The Lord of the Rings: Return of the King* (2003).

In this kind of sequence, there is no absolute bearing, but there is a more general sense of common bearing. We may analyze this as a relation on scene graphs that requires two leylines to have the same slope and sign (like absolute bearing), but not the same anchor points (unlike absolute bearing). Thus film space like this inhabits a systematic abstraction of AB-coherence.

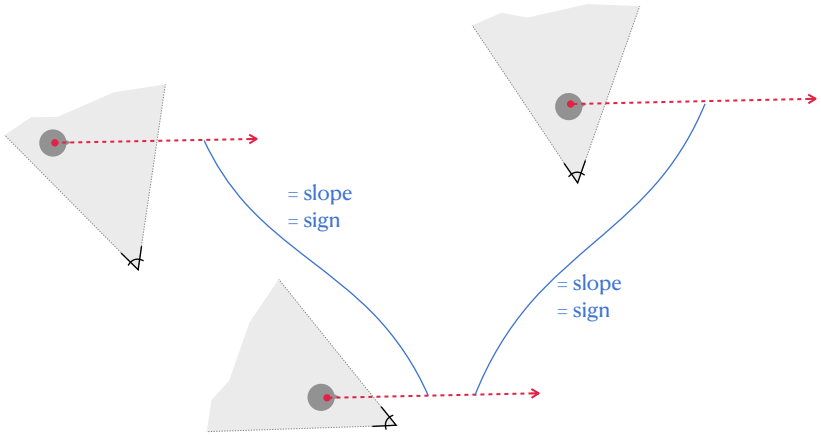

Figure 16: Scene graph for alignment editing: leylines in each shot are encoded as as parallel, but otherwise disconnected.

I expect that other forms of partial spatial coherence will be found in film practice, that likewise play with relations of anchor, slope, and sign. There may also be forms of spatial coherence that exploit paths, or common destinations, rather than straight lines, to hold the scene graph together. I flag these questions for future research.

4 Viewpoint constraints and coherence

If each shot contributes a single view space, how is absolute bearing— a relation between view spaces— established for an entire scene? One solution we've already encountered is the use of master shots to explicitly depict relations of absolute bearing. But without such summary information, contextual knowledge of the linear relations between shots per force plays a larger role, and a more targeted strategy is called for. In this section, I'll argue that a system of conventional constraints on the evolution of viewpoint, now ubiquitous in mainstream film, is one of the central tools that filmmakers and viewers exploits to secure spatial coherence.

4.1 Sight lines, action lines, and leylines

Wherever there are straight lines in a film's action that intersect more than one view space, there is the source material for establishing absolute bearing. Linear objects (like roads, walls, or pathways) or linear actions (like glancing, shooting, pointing, or following) that extend from one shot into another are well suited to ground the construction of AB-coherent leylines. This is part of why sight lines and action lines play such a central role in film making practice.

Consider point-of-view (POV) editing: a typical POV sequence, like that of Fig. 4 from *Vertigo*, consists of a **glance shot** of character looking at something, followed by an **object shot** of the thing

they are looking at, from their point of view. Interpreting a sequence as POV requires that the viewpoint of the object shot coincide with the position or the eyes of the character in the glance shot.[25] Since the glance specifies an implicit **line of sight**, and since the object shot locates this line at its center, a sight line is understood to run through both shots. Its position and direction is depicted in the glance shot, and fixed by convention in the object shot. But once these directions are established, absolute bearing follows directly. Thus POV editing provides a fast track to absolute bearing, and a powerful spatial tool for the film editor.

Of course, sight lines in the service of POV are just one kind of spatial link on the way to absolute bearing. The more general class of **action lines** are those linear relationships implied by any kins of spatially directed action. What is important, in general, is that an action line be depicted in one shot, and continue either explicitly or implicitly into the next shot. By showing, shifting, and managing action lines, a filmmaker can secure the spatial cues required to build a lattice of AB-coherent view spaces. Animator Craig Good described such a process this way:[26]

> On *A Bug's Life* [1998] we had painted ourselves into the lack of a corner by designing the ant bunker as an ovoid, featureless cave. It had almost no geography to orient the viewer, so we had to carefully place the stage line (what we called "the line" from the 180° rule) and move it deliberately, using the characters themselves as geography. Notice the way Hopper paces back and forth. Not one of those steps is random: They are all moving the line to where we needed it to be.

When a pair of shots show the beginning and end of a linear action, then we have the position and orientation of both ends of an action line, so absolute bearing follows. So, for example, if a stick is thrown in one shot, and the stick is caught in the next shot, then we may infer (i) a linear relationship between the throw and the catch; (ii) the direction of the throw in the initial shot; (iii) the direction of the catch in the final shot. In this case, the interpretation of the shots alone, together with the world knowledge necessary to identify the beginning and end of the same action, are sufficient to establish absolute bearing.

But there are a variety of ways that shot content and world knowledge fall short of absolute bearing. When an action line is encoded in one shot, but in the next, its direction is not encoded, then there is not enough information in the scene to establish absolute bearing. Yet is it is just here that system of conventional constraints have arisen bridge the gap between shot contents and scene graphs with absolute bearing.

[25] See Cumming, Greenberg, Kaiser, et al. 2021 for details and generalization.

[26] From his 2018 answer to the Quora.com discussion "What's so bad about violating the 180 degree line?" Url: https://qr.ae/pvoaKo

4.2 The XTR-System

Recent work by Cumming, Greenberg, Kelly and colleagues (2017; 2021; 2021) (henceforth "CGK"), documents a family of conventional **viewpoint constraints** at work in mainstream film. Such constraints govern the dynamics of viewpoint across sequences, in relation to a central line of action. They include the X-Constraint, already well-known as the 180° rule, as well as the T-Constraint, R-Constraint, and others.[27] CGK argue that the so-called XTR-System is ubiquitous in film and television, and informs spatial interpretation as much as film-making practice.

Although I will follow CGK in describing X, T, and R as viewpoint constraints, they can just as well be thought of as *action-line constraints*: given an action line, how is its position in space fixed across two shots? Understood this way, the XTR-System clearly contributes to the project, anticipated above, of leveraging action lines to construct AB-coherent leylines. In this section, I'll review the XTR-System, with special focus on the X-Constraint, showing how each constraint triangulates action lines and viewpoints to bring about spatial coherence through absolute bearing.

4.2.1 The X-Constraint

The X-Constraint requires that adjacent shots preserve the screen direction of a salient line of action, where screen-direction is understood as overall orientation along the X-axis of the shot space (i.e. left-right direction on the screen).[28] To illustrate, compare the following pair of sequences, and assume that the ball depicted travels in a straight line, as indicated by the arrow, through the whole scene. The sequence in Figure 17 preserves screen direction along the X-axis, so conforms with the X-Constraint, while that in Figure 18 is incompatible with the X-Constraint.

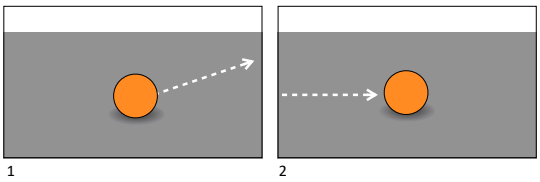

Figure 17: X-Constraint conforming.

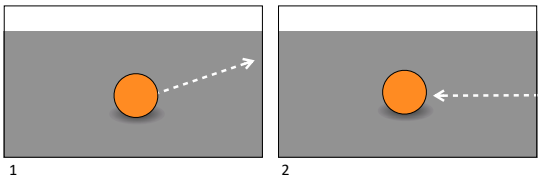

Figure 18: X-Constraint incompatible.

The X-Constraint is standardly operationalized as requiring that successive camera positions stick to the same side of the action line, illustrated in Figure 19. As the reader may verify, camera positions on opposite sides of the action line yield contrasting screen directions for the action line itself.[29]

[27] Point-of-view and related editing techniques belong to a separate system of viewpoint constraints that center on the perceptual experience of depicted characters.

[28] For discussion and references, see Bordwell, Thompson, and Smith 2017, pp. 231–33, Huff and Schwan 2012, Cumming, Greenberg, and Kelly 2017, and Levin and Baker 2017.

[29] Cumming, Greenberg, and Kelly 2017, p. 13.

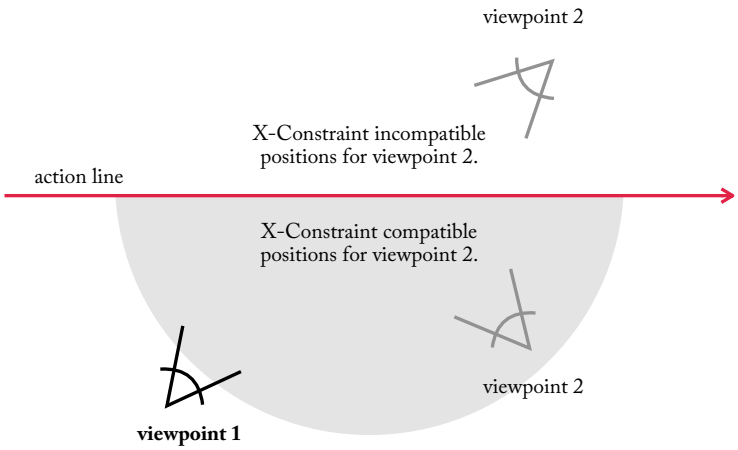

Figure 19: The X-Constraint implies a constraint on camera positions: "don't cross the line."

CGK (2017) analyze the X-Constraint as a relation between viewpoints, and situate it within a broader class of conventional viewpoint constraints. As they point out, screen direction itself is a function of the spatial relationship between the action line and the viewpoint of the current view space. Screen direction, as it figures in the X-Constraint, can be precisified as **X-direction**: the sign (positive, negative, or null) of the action line as projected to the X-axis of the coordinate frame of the viewpoint. Two X-directions are **consistent** just in case they don't have opposite signs. Then the X-Constraint can be stated as follows:

(4) **X-Constraint**:
If two shots A, B are connected by the X-constraint, then,
given a view space v_A for A and v_B for B:

a. view space v_A includes part of an action line α;

b. view space v_B includes part of α;

c. the X-direction of α in v_A is consistent with the X-direction of α in v_B.

CGK make the case that the X-Constraint is more than rule of thumb for filmmakers, or an empirical generalization about filmmaking. Instead, viewers actually assume the X-Constraint in their interpretation of scenes. To draw out this effect, CGK invite us to consider a scene of two men playing chess, where the players and the chess board never appear in the same shot, yet we are still left with a clear intuition about their relative position. Here is a representative pair of shots from the scene.

1 2

Figure 20: The Chess Case. Is the blond man playing white or black?

The two shots alone are optically compatible with two different interpretations: that the blond is playing black and that the blond is playing white. Nor does world knowledge settle the question of who is playing what. Yet we clearly infer that the blond is playing black. The natural explanation of this is that we assume the X-Constraint: if screen direction is preserved, then it follows which player is on which side of the board.

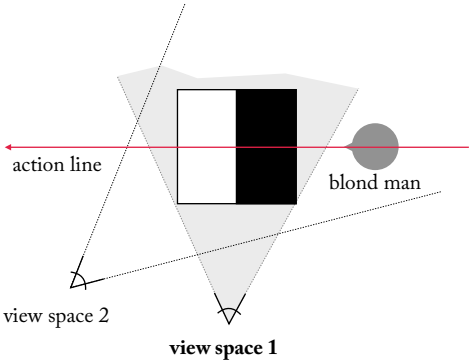

Figure 21: Interpretation compatible with the X-Constraint.

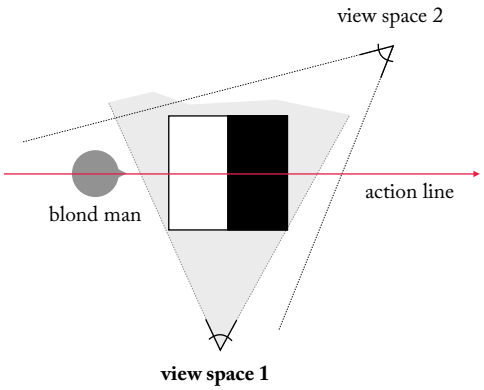

Figure 22: Interpretation incompatible with the X-Constraint.

Examples like the Chess Case suggest that the X-Constraint "reflects a form of default spatial encoding" (Levin and Baker 2017, p. 11): absent explicit cues to contrary, viewers assume the X-Constraint, and filmmakers depend on viewers to assume this. But *why* should the X-Constraint be the default? Why are we not content to leave the Chess Case spatially unresolved? After all, we've seen that viewers are unperturbed other kinds of spatial uncertainty.

I believe the answer to these questions is that viewers and filmmakers prize absolute bearing, and the X-Constraint is a means to this end. The X-Constraint on it's own doesn't achieve this end: allowing that two ends of an action line have the same X-direction doesn't fix their overall direction with respect to one another, the relation required for absolute bearing. Still, we may hypothesize that the promotion of absolute bearing is the primary function of the X-Constraint. When a single

action line is present across shots, the X-Constraint works to narrow the possible ways such a line can pass through two the view spaces. Since absolute bearing itself is a property of a line between view spaces, the X-Constraint feeds directly into the stream of information necessary to get to full spatial coherence.

Consider how the X-Constraint serves absolute bearing in the Chess Case. The shot of the chess board depicts the axis of play, so encodes the direction of a leyline ℓ anchored on the chess board. World knowledge assures us that the axis of play in the second shot is the same as that in the first. From this, we know that ℓ is anchored to the blond man, and that ℓ runs parallel to the line that passes through the center of the two player's heads. In other words, we know the slope of ℓ in view space 2 as well. But we don't know the sign. Thus, world knowledge, visual cues, and shot content don't allow us to construct an AB-coherent scene graph. Instead, we are left with two possible coherent interpretations. Only when the X-Constraint is added can AB-coherence be established.

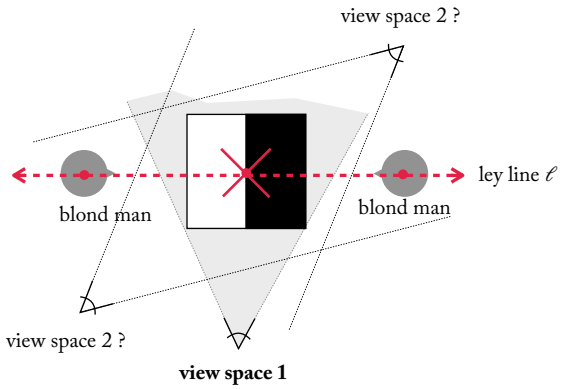

Figure 23: Without the X-Constraint, the scene graph for the Chess Case establishes the slope and anchor of the players relative to the chess board, but not the sign.

The case exemplifies a general pattern. The X-Constraint is always used in conjunction with a range of other cues to establish absolute bearing. When these cues fall short, the X-Constraint works as an ancillary assumption on the way to full spatial coherence. The central force of the X-Constraint in these situations is to secure the sign of the relevant action line, when its anchor points and slope are already known.

4.2.2 The T-Constraint

Whereas the X-Constraint requires that the X-direction of the action line be preserved between shots, T and R impose more stringent demands. The effect of the T-Constraint is to require that the action line have the same (overall) screen direction and screen position across shots. The following

pair of cases illustrates the idea.

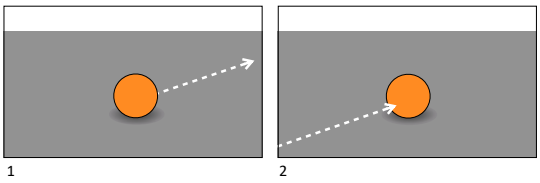

Figure 24: T-Constraint conforming.

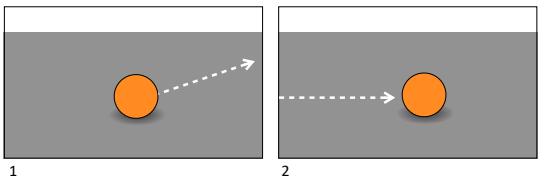

Figure 25: T-Constraint incompatible.

Evidence for the interpretive efficacy of the T-Constraint comes from the sequences in Fig. 30, due to CGK (2017, pp. 16–19), in which a cue ball is struck and rolls out of frame in the first shot, and its entrance into the second shot is anticipated. The white arrows show the actual path of motion (in the first shot) and the anticipated path of motion (in the second shot). The annotations in the second image above show two possible interpretations of the case (but are not part of the original sequence).

Figure 26: The Pool Case. Will the cue ball enter the second shot at 45 degrees or 10 degrees?

Considered in isolation, the contents of the two shots are compatible with innumerably many interpretations, including the two highlighted here. Of course, world knowledge does not narrow the space of options. Yet CGK report that viewers overwhelmingly favor 45° over 10°. Note that the preferred interpretation cannot be derived by the X-Constraint alone, because X is compatible with both. But when the T-Constraint is assumed, the 45° interpretation follows directly and the 10° interpretation is ruled out.

CGK analyze the T-Constraint, like the X-Constraint, as a viewpoint constraint. In the case of T, it requires that the position of the second viewpoint be derived from that of the first by translation, without rotation, along the vector of the action line:

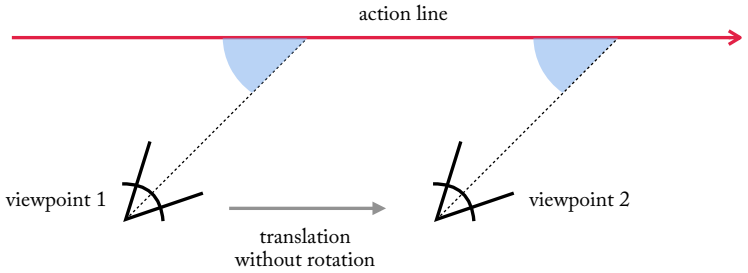

Figure 27: The T-Constraint requires that subsequent viewpoints be related by translation without rotation parallel to the action line

Like the X-Constraint, the T-Constraint can be defined as a relation between view spaces relative to an action line that connects them:

(5) **T-Constraint**:
If two shots A, B are connected by the X-constraint, then,
given a view space v_A for A and v_B for B:

a. view space v_A includes part of an action line α;

b. view space v_B includes part of α;

c. the direction of α in v_A = the direction of α in v_B.

Unlike the X-Constraint, the T-Constraint doesn't merely put a *constraint* on the direction of the action line in the second shot relative to the first, it *determines* the direction of the action line in the second shot relative to the first. This is an important property, because it means that, so long as the direction of the action is fixed by the initial shot, then applying the T-Constraint is one way of establishing absolute bearing. Indeed, the Pool Case is a prototypical example of absolute bearing: one can imagine entering into the space of each shot and knowing how to point to the pool balls (or cue ball) located in the other. Thus, thanks to the T-Constraint we have clear judgements about the *direction* the cue ball will enter the second view space, even if we have little exact sense of the *distance* it must travel to get there.

4.2.3 The R-Constraint

Whereas the T-Constraint requires that the overall screen direction of the action line stay constant between shots, the R-Constraint requires that it be *reflected* across the central axis of the view space in the two shots, as illustrated below.

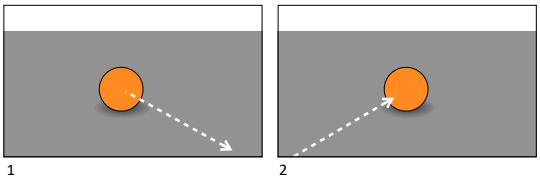

Figure 28: R-Constraint conforming.

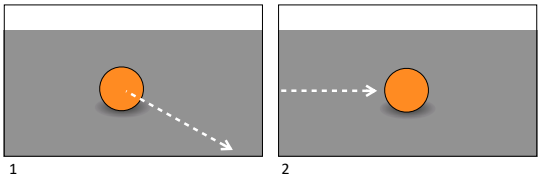

Figure 29: R-Constraint incompatible.

To implement this idea as a constraint on viewpoint, CGK (2021, pp. 743–45) define the R-Constraint in terms of the matched angle of the viewpoints in the two shots relative to the action line. We may state the constraint equivalently in terms of reflection across an axis normal to the action line. The R-Constraint is incompatible with the T-Constraint, but likewise entails the X-Constraint. (Since the reflection is across the normal to α, not α itself, X-direction is always preserved.)

(6) **R-Constraint**:
If two shots A, B are connected by the R-constraint, then,
given a view space v_A for A and v_B for B:

a. view space v_A includes part of an action line α;

b. view space v_B includes part of α;

c. the direction of α in v_A is the reflection the direction of α in v_B (across the normal of α).

This definition may be visualized as follows:

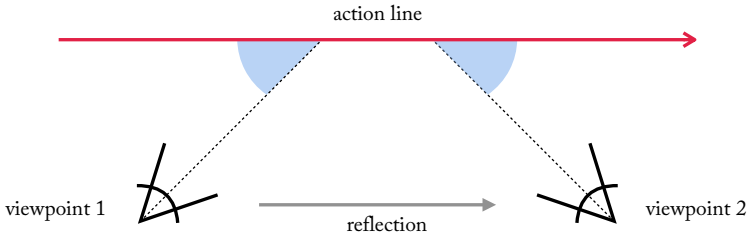

Figure 30: The R-Constraint requires that subsequent viewpoints be related by reflection across the normal to the action line.

The R-Constraint is like the T-Constraint, and unlike the X-Constraint, in that it determines the position and direction of the action line across two shots, rather than merely constraining their relation. At the same time, since R and T are mutually incompatible, applying the R-Constraint amounts to a distinct way of establishing absolute bearing from T. This is apparent in the sequence from Fig. 28, where relations of direction, but not relations of distance, are fixed between the two shots, in manner that is now familiar.

4.3 Context, constraints, and coherence

My conjecture is that viewpoint constraints in the XTR-system function to help viewers and filmmakers establish absolute bearing. They acts as ancillary assumptions which reinforce displayed cues, or fill the gap when cues are absent. The interaction of viewpoint constraints and context to achieve coherence is complex but systematic. On one hand, T and R impose very strong spatial requirements, but are applied only rarely. On the other, X contributes relatively weak demands, but is almost constantly in use. These facts reflect a general pattern: when context contains rich spatial information, the viewpoint constraints play little or no role in establishing absolute bearing; but as the information supplied by context decreases, the role of viewpoint constraints grows. The interaction of context with AB-coherence helps explain why some sequences flout the viewpoint constraints without incoherence, while others automatically imply viewpoint constraints even without explicit cues of conformity.

To map out these effects, suppose we have a sequence of two shots, with the first depicting the initiation of an action line, and the second depicting a space where, given the context, we expect the action line to extend. Since the first shot explicitly depicts the beginning of the action line, it encodes that line's anchor, slope, and sign. Such a shot might depict a rolling ball, a stick being thrown, a pointed finger, or a gaze.

We can classify sequences like this as high, middle or low information conditions, depending on the amount of information contributed by the second shot. In a **high information condition** the anchor, slope, and sign of the action line are all encoded in the second shot. In a **middle information condition** only the anchor and slope of the action line is encoded. In a **low information condition** only the anchor is defined.

In all of these conditions, we can say that the **basic scene**— that is, the shot contents, together with contextual reasoning and world knowledge, but without viewpoint constraints— fixes the identity of the action line in the two shots. The question we may then ask is: which viewpoint constraints must be added to the basic scene in order to establish absolute bearing? A summary of the answers discussed below is provided in Fig. 38.

4.3.1 High information conditions

In high information conditions, both shots in the sequence encode the anchor, slope, and sign of an action line and the basic scene encodes that they are two ends of the same line. A sequence showing a ball rolling out of frame in one shot, and then rolling into frame in the next satisfies this description.

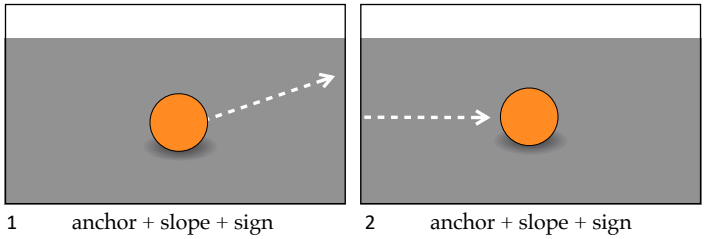

Figure 31: High information condition: anchor, slope, and sign are all explicitly depicted in the second shot.

In the high information condition, the basic scene alone establishes absolute bearing. As a consequence, X, T, R are all optional with respect to AB-coherence. High information sequences may conform with any one of these constraints; but they may also be incompatible with X, T, or R and still be AB-coherent. Thus the same sequence from Fig. 31 conforms with X, for example, but is incompatible with with T or R.

It's well-known among film scholars that the X-Constraint does not apply in all situations.[30] There are of course cases, like the sequence from *Breathless*, where the X-Constraint is expected, but not met in an expected way, and spatial incoherence results. But there are also many examples of so-called " 180° violations" where the X-Constraint does not apply, but spatial coherence is maintained nonetheless (indeed, the violation often goes unnoticed[31]). Here I include a constructed example, and an example from *Stagecoach* (1939), both following the same visual logic.

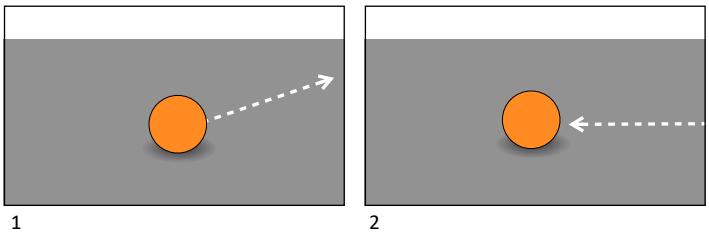

Figure 32: A constructed X-Constraint incompatible sequence.

[30] See e.g. Carroll 2008, pp. 119–20, Bordwell, Thompson, and Smith 2017, pp. 239–40, Cumming, Greenberg, and Kelly 2017, pp. 21–24; all discuss the example from *Stagecoach* below.
[31] See Levin and Baker 2017, pp. 9–10.

1 2

Figure 33: An X-Constraint incompatible sequence from *Stagecoach* (1939).

These sequences are spatially coherent despite incompatibility with the X-Constraint because they encode high-information conditions: anchor, slope, and sign are encoded in both shots, so AB-coherence holds. In such cases, I propose, even the X-Constraint is optional.

4.3.2 Middle information conditions

In middle information conditions, the second shot encodes only the anchor and slope of the action line, while the basic scene alone encodes that they are two ends of the same line. For example: a sequence showing a ball rolling down a road in one shot, and then a shot of a stretch of empty road, where we assume that a single road is depicted throughout, and that the ball will eventually enter the view space of the second shot. The second shot depicts the anchor and slope of the action line, but not the sign: i.e. it doesn't depict which direction the ball is coming from.

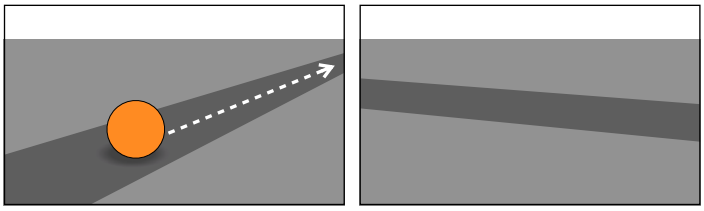

Figure 34: Middle information condition: anchor and slope, but not sign, are explicitly depicted in the second shot.

In middle information sequences, the basic scene alone doesn't encode enough information to secure absolute bearing. But, when the X-Constraint is added, the scene becomes AB-coherent. As we saw in Section 4.2.1, the Chess Case was a middle information condition just like this. The anchor points were the chess board and the blond man. The action line (from white to black, say) was the axis of play. We are given the slope of the axis of play in both shots. And the first shot

fixes the sign of this line, but the second shot does not. As we saw above, the X-constraint played a key role in finally determining the sign of the action line in the second shot.

Figure 35: The Chess Case as a middle-information condition: the anchor, slope and sign of the action line (white arrow) are encoded in the first shot; but only the anchor and slope are encoded in the second shot.

In general, middle information conditions require the X-Constraint for AB-coherence; given X, however, T and R are optional. Thus they may be flouted without any incoherence, as they are in the Chess Case.

4.3.3 Low information conditions

In low information conditions, the second shot encodes the anchor point of an action line from the first shot, but nothing else. This kind of configuration arises when the second shot reveals an obvious end point for a linear action, but not its path. To get this result in the following example, assume that the motion of the orange ball is aimed at the white ball, and that they will eventually collide. Here we know neither the slope or the sign of the action line from the basic scene alone, since it isn't encoded where, or in what direction, the orange ball will enter the view space of the second shot.

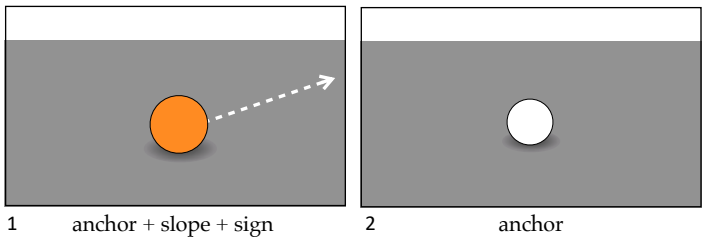

Figure 36: Low information condition: anchor, but not the slope or sign, are explicitly depicted in the second shot.

In low information conditions, of course, the basic scene alone does not establish absolute bearing. Even with the addition of the X-Constraint, absolute bearing does not follow. Instead,

either T or R are necessary in order to establish AB-coherence. Since T and R are mutually incompatible, only one can be applied at a time, but at least one must be applied.

The Pool Case was a low information condition exactly like this. The first shot shows the pool stick striking the cue ball, so encodes the anchor, slope, and sign of the line of action. The second shot depicts its expected target(s), but encodes neither the slope nor sign of the anticipated action line. CGK's forced-choice judgement between 45° and 10° interpretations (see Section 4.2.2) favored a a T-Constraint reading. But viewers may also feel the pull of an R-Constraint reading, where the cue ball enters at a 45° angle from the top of the screen, to strike the orange 5-ball.

1 2

Figure 37: The Pool Case as a low-information condition: the anchor, slope and sign of the action line (white arrow) are encoded in the first shot; but only the approximate anchor point (white circle) is encoded in the second shot.

These considerations demonstrate the power of T and R with respect to absolute bearing, creating coherence out of nearly thin air. But they also explain why we see the most vivid demonstrations of T and R in especially low-information settings, like a shot of a motionless pool ball on a featureless plane.

4.3.4 Summary

We may summarize the findings of this section with the following table. It shows two broad trends. First, there is a trade-off between the information available in shot 2 and the strength of the viewpoint constraint that must be assumed in order to establish AB-coherence. And second, as the information available in shot 2 increases, a greater number constraints can be flouted while maintaining AB-coherence. These are telling indicators of when, and why, viewpoint constraints are deployed in film— the topic of the next section.

	shot 1	shot 2	conformity required for AB	incompatibility consistent w/ AB
high information condition	anchor slope sign	anchor slope sign	∅	X, T, R
middle information condition	anchor slope sign	anchor slope	X	T, R
low information condition	anchor slope sign	anchor	T, R	T (not R) R (not T)

Figure 38: The interaction of contextual information, viewpoint constraints, and absolute bearing.

4.4 Maximize coherence

Different viewpoint constraints seem to be present in some sequences and not others, required in some conditions but optional in others. What general principles govern when a given viewpoint constraint will or will not apply?

CGK (2017, pp. 22–23) point out that viewpoint constraints are never expected to apply when doing so would lead to absurdity. We don't assume the X-Constraint in the *Stagecoach* sequence, or the R-Constraint in the Chess Case, because these assumptions would directly conflict with information already encoded in the shots as supplemented by narrative and world knowledge. This establishes an upper limit on constraint application. But why then do viewpoint constraints apply at all? Why assume X in the Chess Case, in the first place? And why assume T in the Pool Case?

The findings of Section 4.3 demonstrated that viewpoint constraints are often the essential bridge between context and coherence; the less information supplied by context, the more work viewpoint constraints must do. Assuming that AB-coherence is the goal of filmmakers and viewers alike, this pattern suggests a default maximization: we should apply as many constraints as context will allow, so as to secure absolute bearing as often as possible. We may articulate such principle this way:[32]

(7) **Maximize Coherence**
When a scene encodes that the same action line intersects two shots, assume as many constraints from the set $\{X, T, R\}$ as possible without conflicting with other information encoded by shot contents, world knowledge, and narrative context.

[32]Compare to the *Maximize Discourse Coherence* principle from Asher and Lascarides (2003), within the SDRT framework.

Maximize Coherence implies that for each constraint, so long as the basic scene already *conforms* with a constraint, it will apply; and so long as the basic scene is *incompatible* with the constraint, it will not apply. But in addition, if the basic scene is merely *compatible* with a constraint, but doesn't explicitly conform to it, then the constraint must still apply. Thus, in a low information setting, the principle requires that we assume either T or R. (The issue of whether it is T or R that should be applied is not resolved here, but remains an interesting open question.) In a middle information setting, Maximize Coherence requires that we assume at least X. And in a high information setting, it doesn't require any constraint. Thus Maximize Coherence functions to facilitate absolute bearing at every turn.

An important consequence of Maximize Coherence is that constraints are applied when the information encoded in a basic scene already conforms with these constraints. In an X-Constraint conforming sequence, like our initial cases from *Totoro* or *Sesame Street*, there is enough information in the basic scene to establish absolute bearing, but the scene also conforms with the X-Constraint; according to Maximize Coherence, the X-Constraint applies here too. Such information is redundant from a purely logical perspective, but not from a processing perspective. Filmmakers constantly make use reinforcing cues, precisely because film sequences pass so quickly and tend to be packed with motion and narrative (unlike the Chess and Pool cases!). The more cues that point to the same spatial resolution, the better. Since most scenes are at least *compatible* with the X-Constraint, this means the X-Constraint is constantly in use in mainstream film.

Maximize Coherence is a generalization of the idea that the X-Constraint is a default of spatial coherence, as envisioned recently by Huff and Schwan (2012), Cumming, Greenberg, and Kelly (2017, pp. 23–24), Levin and Baker (2017, pp. 9–11), among others. In the generalized version, it's not just X, but T and R that are defaults of spatial interpretation. We can now understand this default not just as a primitive fact about the interpretive mechanism, but as a strategic deployment of assumptions to maximize the chance of establishing absolute bearing.

5 Conclusion

I began this essay by casting film interpretation as a dynamic process of incremental update to a central discourse record. For the spatial content of a film, I suggested that the discourse record takes the form of a scene graph, a set of view spaces, contributed by individual shots, connected by a lattice of straight lines. Scene graphs are only fully resolved, or coherent, when absolute bearing is established for these linking lines, a spatial standard looser than full metric coherence. Eschewing requirements on the representation of distance, absolute bearing instead implies that every shot depicts at least one object from which we can compute the bearing of another object in another shot, and vice versa. Given that incoherent film spaces are difficult to understand and to remember, it is plausible that scene graphs with absolute bearing reflect the mind's native coding

of extended space. My final proposal was that the XTR-System of viewpoint constraints is a set of conventional mechanisms that help leverage depicted action lines into relations of absolute bearing. Viewpoint constraints work in lock-step with contextual information to produce stable and coherent film spaces.

Film itself is microcosm of lived experience. It offers up to its viewers a sampling of information from across the domains of life outside of film: representations of events, narrative, time, space, sound, emotion, other minds, and social relations are all engaged. Film scholarship has traditionally focused on the low-level role of visual perception, on one hand, and the high-level role of narrative comprehension, on the other, as the chief drivers of interpretation. If the argument of this essay is right, then cognitive maps also play a critical role. Indeed, perhaps the most celebrated feature of film— the ability to create coherent scenes from discontinuously edited cuts— seems to turn centrally on the strategic deployment of mental maps.

It seems there is more to film interpretation than merely recovering a body of information made available by the filmmaker. It is also a matter of engaging the relevant psychological capacities in precise coordination with those intended by the film's production. Visual perception, cognitive mapping, event segmentation, theory of mind, and more, come in and out of phase as the film unfolds. In this way, films become signifying machines designed to orchestrate the whole of human cognition.

References

Abusch, D. (2012). "Applying discourse semantics and pragmatics to co-reference in picture sequences". In: *Proceedings of Sinn und Bedeutung 17*.

Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Bateman, J. A. and K.-H. Schmidt (2012). *Multimodal Film Analysis: How Films Mean*. Vol. 5. Routledge.

Beatty, J. (2004). *Film Glossary*. URL: `http://userhome.brooklyn.cuny.edu/anthro/jbeatty/COURSES/glossary.htm` (visited on 2004).

Berliner, T. and D. J. Cohen (2011). "The illusion of continuity: Active perception and the classical editing system". In: *Journal of Film and Video* 63.1, pp. 44–63.

Bordwell, D. (2002). "Intensified Continuity Visual Style in Contemporary American Film". In: *Film Quarterly* 55.3, pp. 16–28.

Bordwell, D. and K. Thompson (2010). *Film Art, 9th Edition*. New York, NY: McGraw-Hill.

Bordwell, D., K. Thompson, and J. Smith (2017). *Film art: An Introduction, 11th edition*. McGraw-Hill New York.

Bordwell, D. (1985). *Narration in the fiction film*. University of Wisconsin Press.

Burch, N. (1981). *Theory of film practice*. Princeton University Press.

Carroll, N. (2008). *The Philosophy of Motion Pictures*. Wiley-Blackwell.

Cohn, N. (2018). "In defense of a grammar in the visual language of comics". In: *Journal of Pragmatics* 127, pp. 1–19.

Cumming, S., G. Greenberg, and R. Kelly (2021). "Temporal Continuity". Video essay.

Cumming Samuel, S., G. Greenberg, E. Kaiser, et al. (2021). "Showing Seeing in Film". In: *Ergo* 7.
Cumming, S., G. Greenberg, and R. Kelly (2017). "Conventions of viewpoint coherence in film". In: *Philosopher's Imprint* 17.1.
Cutting, J. E. and A. Candan (2013). "Movies, evolution, and mind: From fragmentation to continuity". In: *The Evolutionary Review* 4.3, pp. 25–35.
Epstein, R. A. et al. (2017). "The cognitive map in humans: spatial navigation and beyond". In: *Nature neuroscience* 20.11, pp. 1504–1513.
Frith, U. and J. E. Robson (1975). "Perceiving the language of films". In: *Perception* 4.1, pp. 97–103.
Gallistel, C. R. (1990). *The organization of learning.* The MIT Press.
Gernsbacher, M. A. (1997). "Coherence cues mapping during comprehension". In: *Processing interclausal relationships. Studies in the production and comprehension of text*, pp. 3–22.
Greenberg, G. (2020). "The Structure of Visual Content". Manuscript.
Heim, I. (1983). "File change semantics and the familiarity theory of definiteness". In: *Meaning, Use, and Interpretation of Language*. Ed. by R. Bäuerle, C. Schwarze, and A. von Stechow. Berlin: de Gruyter, pp. 164–189.
Hobbs, J. R. (1985). *On the Coherence and Structure of Discourse*. Center for the Study of Language and Information.
Huff, M. and S. Schwan (2012). "Do not cross the line: Heuristic spatial updating in dynamic scenes". In: *Psychonomic Bulletin & Review* 19.6.
Kamp, H. (1981). "A theory of truth and semantic representation". In: *Formal semantics-the essential readings*, pp. 189–222.
Katz, E. and R. D. Nolan (2012). "The Film Encyclopedia". In:
Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Center for the Study of Language and Information.
Kraft, R. N. (1987). "Rules and strategies of visual narratives". In: *Perceptual and Motor Skills* 64.1, pp. 3–14.
Kraft, R. N., P. Cantor, and C. Gottdiener (1991). "The coherence of visual narratives". In: *Communication Research* 18.5, pp. 601–616.
Levin, D. T. and C. Wang (2009). "Spatial Representation in Cognitive Science and Film". In: *Projections* 3.1, pp. 24–52.
Levin, D. T. and L. J. Baker (2017). "Bridging views in cinema: A review of the art and science of view integration". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 8.5, e1436.
Lewis, D. (1979). "Scorekeeping in a Language Game". In: *Journal of Philosophical Logic* 8, pp. 339–59.
Loschky, L. C. et al. (2020). "The scene perception & event comprehension theory (SPECT) applied to visual narratives". In: *Topics in cognitive science* 12.1, pp. 311–351.
Maier, E. and S. Bimpikou (2019). "Shifting perspectives in pictorial narratives". In: *Proceedings of Sinn und Bedeutung*. Vol. 23. 2, pp. 91–106.
McCloud, S. (1993). *Understanding Comics: The Invisible Art*. Kitchen Sink Press.
Meilinger, T. (2008). "The network of reference frames theory: A synthesis of graphs and cognitive maps". In: *International conference on spatial cognition*. Springer, pp. 344–360.
Peer, M. et al. (2021). "Structuring knowledge with cognitive maps and cognitive graphs". In: *Trends in cognitive sciences* 25.1, pp. 37–54.
Shaviro, S. (2016). "Post-continuity: an introduction". In: *Post-Cinema: Theorizing 21st-Century Film*. Ed. by S. Denson and J. Leyda. Reframe Books.

Stalnaker, R. (1978). "Assertion". In: *Syntax and Semantics, Volume 9: Pragmatics*. Ed. by P. Cole and J. Morgan. Academic Press, pp. 315–332.
Stork, M. (2011). *Chaos Cinema*. Press Play.
Tan, E. S. (2018). "A psychology of the film". In: *Palgrave Communications* 4.1, pp. 1–20.
Wildfeuer, J. (2014). *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*. Vol. 9. Routledge.
Wildfeuer, J. and J. Bateman (2016). "Linguistically oriented comics research in Germany". In: *The visual narrative reader*, pp. 19–66.