Philosophers' Imprint

VOLUME 000, NO. 000

JANUARY 0000

CONVENTIONS OF VIEWPOINT COHERENCE IN FILM

Samuel Cumming[*]

Gabriel Greenberg[*]

Rory Kelly[†1]

[*]*Department of Philosophy*
University of California, Los Angeles

[†]*Department of Film, Television and Digital Media*
University of California, Los Angeles

1. All authors contributed equally to this work. The film sequences were produced with assistance of Judy Phu (Cinematography), Jean Tyson (Cinematography), Dylan Chapgier (Cinematography) , Nathan Mead (Assistant Camera/Gaffer), David Preggerson (Production Assistant), Jonathan Oldham (Actor), and David Atlas (Actor). We are grateful for the comments and suggestions of reviewers and colleagues, as well as audiences at the University of Oslo, Cornell, SCSMI, UCLA, and Haverford. This project was supported in part by the UCLA Office of the Vice Chancellor for Research Transdisciplinary Seed Grants.

Over the last century, the philosophical study of meaning, including the areas of semantics and pragmatics, has overwhelmingly focused on *language*. In this paper, we examine the interplay of semantics and pragmatics in the alternative domain of *film*— understood to include all variety of video and animation, as they appear in cinema, television, and online. Far from a marginal case, film is one of the primary means by which humans exchange narrative in contemporary society.

Films consist of one or more shots organized into linear temporal sequence; our focus are those films which combine multiple shots to express their content. While, in principle, each shot in a given sequence could depict a completely isolated situation with no implied connection to the next, filmmakers have discovered methods of *montage*, or editing, by which shots are arranged so as to reliably convey unbroken stories, taking place in continuous space and time. How sequences of distinct images (and sounds) are able to express such coherent content is one of the central questions of classical film theory, and is the subject of this paper.[2]

The SEMANTIC VIEW of film holds that the expression of unified content through the juxtaposition of shots is made possible in part by a system of semantic conventions. Like the rules of language, such conventions constitute specific interpretive principles which go beyond our everyday capacities for perception, thought, and action. In the case of film, they function to introduce implicit temporal, spatial, and narrative relationships between the events represented by adjacent shots. But, unlike some of the examples familiar from language, these conventions need not be fully arbitrary; instead, they may be naturally motivated in part by prior expectations and tendencies. For example, edited sequences following the common *point-of-view* convention move from a shot of a character's face to a shot of the scene they perceive, echoing the natural transition from noticing a glance to looking at the

2. Of course, editing is just one of the mechanism that can drive film narrative. Filmmakers may exploit techniques of camera movement and mise-en-scène to convey in a continuous shot what could also be approximated by editing.

object of that glance. Versions of the semantic view have been widely explored from a semiotic perspective, most prominently in the work of Metz (1974, chs 3-5).[3] More recent advocates of the semantic view include Bordwell (2008, ch. 2), Bateman and Schmidt (2012), and Wildfeuer (2014).[4]

While scholarship in the semantic tradition has made a certain amount of headway, it also faces a number of obstacles. For example, there is little agreement about the first-order issue of which rules in particular are the ones which govern film representation, and what their exact content is. And the foundations of the semantic view are even more unsure: in what sense of "convention" are the rules of film conventional? how are they analogous to linguistic conventions? and how are they compatible with apparent counterexamples? There are even doubts about the motivations for positing semantic conventions of film in the first place. For perhaps all film interpretation can be understood in purely pragmatic terms instead. Clearly, advocates of the semantic view have some explaining to do.

The absence of decisive answers to these questions has engendered reasonable skepticism, leading a number of scholars to develop a PRAGMATIC VIEW of film, in contrast with the semantic view. Authors in this vein include Harman (1977), Wilson (1989, ch. 10), Currie (1993, 1995, ch. 4), Cutting (2005, 22), and Carroll (2008, 116-121). According to the pragmatic view, films are understood in more or less the same way that the experiences of everyday life are understood: through a holistic mixture of perception, rationality, knowledge of the world, and general principles of psychology— but without regular recourse to specialized rules of film interpretation. The pragmatic view doesn't deny that there are conventions which guide film production, such as genre-based conventions governing character types, plot structure, and sound track. What it denies is that conventions play any significant role in deriving the content of a film. Since filmic representation does not systematically depend on conventions, and conventionality is the hallmark of linguistic representation, pragmatic theorists reject any substantive analogy between film and language. In so far as there *are* any interpretive conventions of film, they hold, these are scattered and unsystematic— notable exceptions rather than the norm.[5]

In this paper we set out to develop a new version of the semantic view, one which offers definite answers to the challenges posed above. We articulate the semantic view for a particular pair of conventions that govern spatial relations between *viewpoints*. One such rule is already well known; sometimes called the "180° Rule," we term it the X-CONSTRAINT; to this we add a hitherto unrecorded rule, the T-CONSTRAINT. Both have the effect, in different ways, of limiting the way that viewpoint (or camera position) can shift through space from shot to shot over the course of a film sequence. Using these as exemplars of semantic conventions in film, we'll argue that in this case the semantic view provides the correct account of the means by which films express their content.

Section 1 outlines the general framework of the semantic view, with particular attention to the role it assigns conventions in its account of film expression. Section 2 defines the X-Constraint and argues that we must treat it as a semantic convention. Section 3 introduces the T-Constraint. Section 4 construes both constraints as a relations of viewpoint coherence, analogous to the coherence relations which govern linguistic discourse. Section 5 addresses a well-known challenge for

3. See Nöth (1995, 463-471) for a survey of work on film in the semiotic tradition.

4. McCloud (1993, ch. 3) develops a similar analysis for comics. Though he observes cultural variation in the application of interpretive rules, his emphasis is on their psychological basis, rather than their conventionality.

5. Thus Currie (1995, 135), softening the message of Currie (1993), allows for the possibility of certain specialized conventional rules governing inter-shot relations. For instance: the technique of fading to black to indicate the passage of time may be an isolated semantic convention. To this we might add: the convention of using xylophone scales to indicate the beginning of a dream or fantasy sequences, and the use of a circular mask to indicate seeing through a telescope. But pragmatic theorists like Currie hold that such rules do not constitute any kind of system, and they crucially disagree with the specific cases of semantic conventions defended below.

the semantic view: successful film sequences which seem to "violate" the X-Constraint. Finally, Section 6 provides a semantic analysis of the X-Constraint and sketches a formal implementation. Section 7 is a conclusion.

1. Semantic Conventions in Film

This paper defends the view that the expression of content by film is governed in part by a system of conventional semantic rules. Before taking up a particular account of these rules in the remainder, we turn in this section to describe the central role of convention in our account of filmic representation.

Our focus will be the kind of communication characteristically aimed for in mainstream film, where filmmakers seek to convey coherent and logical stories, and viewers expect them to achieve this standard.[6] A viewer watching such a film faces the challenge of determining, from infinitely many possible interpretations, what content the film expresses. The problem can in turn be analyzed into two, interdependent sub-problems. The first is that of correctly assigning content to each individual SHOT, or continuous moving image.[7] The second is that of correctly assigning content to a SEQUENCE of shots, once the contents of the component shots themselves are fixed. In the terminology of this paper, a sequence is any concatenation of shots, though our focus will be short sequences comprised of two shots only.[8]

The latter is the distinctive problem created by editing. Each individual shot depicts a space, populated with people and objects, over a short span of time. To reconstruct a coherent story, a viewer must effectively stitch together these fragmentary situations, identifying the spatial, temporal, causal, and narrative relationships that connect each one to the next. For successful communication to take place, the filmmaker and viewer must be aligned regarding the assumed relationships between each shot.

There are no in-principle limits to the cognitive resources viewers may draw upon in attempting to solve this problem. A host of broadly pragmatic factors may come into play, including perceptual and cognitive tendencies, knowledge of the world (including knowledge of the filmmaker's own psychology), and knowledge of the world depicted by the film— all marshaled together by "common sense" and rational inference. So long as these faculties are the common possession of both filmmaker and viewer, and so long as their exercise leads to predictable outcomes, they may form the foundation of successful communication.

Still, even well-informed inference can be difficult, slow, and unpredictable. According to the semantic view, the burden of communication is eased through the existence of a set of antecedently established semantic rules, or conventions. These rules are introduced into a film by the expressive intentions and appropriate signaling of the filmmaker (much like the use of an ambiguous word, where a particular meaning is picked out by speaker intentions).[9] For the viewer's part,

6. Thus we set aside the variety of informational exchanges which underwrite many avant-garde and non-narrative film styles. We also exclude much documentary film from our purview, though such works typically do convey coherent narrative.

7. Note that a shot corresponds to a contiguous segment from a single "run of the camera;" shots themselves are presented by exposing viewers to a rapid flicker of FRAMES, which are themselves still images. Our focus is on the relationship between shots, not between frames. In addition, in most contemporary films, shots include both a visual and auditory track. Here we confine our attention to the purely visual aspects of shots.

8. This definition diverges from one common use of the term "sequence" in film scholarship, to mean a series of shots spanning a relatively long arc of action, possibly composed of several scenes. To clarify, "sequences," as we use the term, should be sharply distinguished from SCENES. Following standard usage among filmmakers, a scene is a spatially, temporally, and causally continuous series of events depicted by a film. An INTER-SCENE sequence contains shots which represent events that belong to distinct scenes, while an INTRA-SCENE sequence contains shots which represent events in only one scene. Our focus will be the latter.

9. "Intention" here should be read loosely to denote some goal-directed representation, one which need not contain a representation of the content of the rules in question. A filmmaker may simply be committed to the expression of a particular content (which may be vague and sense-general in certain aspects), without conscious thought about any rule, and this may still be sufficient to ensure that the rule in question applies.

competence with these rules is analogous to a speaker's tacit competence with linguistic rules. Though viewers are typically in no position to verbalize these rules, they are unconsciously internalized and applied in the course of watching a film.

Following research in formal linguistics (e.g. Hobbs 1985; Kehler 2002; Asher and Lascarides 2003), and recent work by Bateman and Schmidt (2012) and Wildfeuer (2014) on film, we suggest that the semantic content assigned by these rules takes the form of COHERENCE RELATIONS: fixed relations between the situations depicted by individual shots which indicate how these situations "cohere" or are connected. Though they are not explicitly marked, these relations implicitly link shots in a sequence. Filmmakers signal their presence, absence, or combination at different points in a sequence through a variety of cues. But since the number of available coherence relations is limited, the cognitive effort required to coordinate on their use is correspondingly curtailed. Such limitations mark an important divergence from the pragmatic view of film interpretation.[10]

Film scholarship in the semantic tradition suggests that there are a host of such rules, likely governing a variety of possible relations between shots, including those of time, space, and action. In this paper we focus on a special class of coherence relations having to do with VIEWPOINT— the visual perspective or vantage point embodied by each shot in a film. The VIEWPOINT CONSTRAINTS explored in this paper impose restrictions on the spatial relationships that may obtain between the viewpoints of adjacent shots, effectively limiting the dynamics of viewpoint over the course of a film sequence.

The coherence relations of film constitute a SEMANTIC aspect of interpretation in the sense that they are governed by conventions which embody specific rules about how a film signal may be mapped to its content. PRAGMATIC aspects of interpretation, by contrast, are not constituted by specific interpretive rules. Instead, they are governed only by very general pragmatic principles, like the laws of logic or rationality, general psychological capacities, like perception or memory, and shared background knowledge of the world. Interpretation is pragmatic to the extent to which it draws solely upon the same kinds of resources which agents use to understand and act in the non-filmic world; it is semantic to the extent that it draws upon specific conventions of interpretation which do not play a role outside of communication. The distinction is clearly not a precise one, but it usefully isolates broadly separable aspects of the interpretive process.[11]

According to the version of the semantic view developed here, the coherence relations governing film are not, on their own, sufficient to determine the content of a film sequence. Rather, film content is derived only through the interaction of semantic and pragmatic factors. Semantic coherence relations simply establish conventional constraints on the content of a film, while other, pragmatic forces flesh these contents out. As a result, the neat division of labor between semantics and pragmatics envisioned by recent theories of language likely does not apply here. In the case of language, there is plausibly a level of recognizable content that can be derived from a sentence by semantic rules and parameterized features of context alone. In the case of film, though one can isolate the contribution of semantic rules, this contribution does not correspond to a significant level of content. We will see evidence of this in Section 2 for the particular case of spatial content.[12]

10. The theory of discourse coherence is taken up in more detail in Section 4.

11. We are deliberately vague about the boundaries of these concepts. Traditionally, pragmatics has been understood to include rationality in the context of cooperative enterprise (as in Grice 1975, chs 2-3), and world knowledge shared by the parties to the communication (as in the "common ground" of Stalnaker 1978). (See Lepore and Stone (2014, chs 1-2) for a critical overview of this approach.) Here we include general psychological capacities as well, notably perception. The unsettled role of perception motivates perhaps a third category of interpretive mechanism, but such taxonomical issues take us too far afield.

12. This view of filmic expression is inspired by Hobbs' (1993) theory of interpretation as abduction, in which the content of a discourse is derived through a holistic process which draws on both pragmatic reasoning and conventional constraints on meaning. According to Hobbs, the content derived through such

Thus, as we see it, the debate between semantic and pragmatic views of film is not about whether film interpretation is entirely semantic or entirely pragmatic, for the centrality of pragmatic forces is a given. Rather, it has to do with whether semantic rules have any significant role to play at all. The pragmatic view holds that, once the content of individual images has been determined, the content of a visual sequence taken as a whole is worked out through pragmatic processes. If there are a few conventions of film interpretation, they are atypical and unsystematic. As Carroll (2008, 119) flatly concludes, there is "no semantics of cinema."

Methodologically, such a pragmatic view must be the default hypothesis, for it aims to explain the same phenomena as the semantic view, but by appeal to cognitive resources which all parties recognize as already in play, and without invoking further specialized conventions. Thus, for every candidate semantic rule, the onus is on the semantic theorist to demonstrate its necessity. In Sections 2 and 3 we will attempt just this. There we'll provide examples of film sequences whose interpretation, we submit, cannot be explained by pragmatic mechanisms alone, but only by appeal to specific, semantic rules.

At the same time, these rules cannot issue from perception, or some other aspect of low-level cognition, for they are not applied automatically in the manner of perceptual rules. Instead, they are selectively applied to certain kinds of shot-to-shot transitions and not to others. The viewpoint constraints we examine, for example, typically apply to *intra*-scene sequences, but only occasionally to *inter*-scene sequences. And, as we discuss in Section 5, there are still other, more complex factors that modulate their applicability. In all cases, the relevant distinctions are entirely narrative, not perceptual. Thus the use of viewpoint constraints must, at root, be influenced by non-perceptual cognitive processes. We propose that they are in fact conventions— conventions which hold in the first instance between individual filmmakers and individual viewers, and as a generalization, between filmmakers as a group and the wider population of viewers.

Crucially, the rules we propose are not merely conventions *among filmmakers*. Such "stylistic" conventions do exist, like the choice to list credits at the end of a film, or to present films using one of the currently favored aspect ratios. But these conventions primarily serve the interests of those involved in film production and distribution, and themselves play little role in interpretation. By contrast, *semantic* conventions arise from the coordinated expectations of *both* filmmakers and viewers. Semantic conventions specifically pertain to the relationship between a film signal and its content, not merely what kind of signal should be created.

To claim that there are semantic conventions of film is not to claim, as some have thought (e.g. Currie 1993; Carroll 2008, 116-121), that film is otherwise language-like, and in particular does not presuppose that there is a *grammar* of film. Syntactic (or grammatical) rules for language provide fixed constraints on how sentences must be structured, on pain of ungrammaticality. But we posit no such rules for film. For coherence relations, when applied, provide constraints only on how a film should be *interpreted*, not on how it should be constructed.[13] If a particular coherence relation is not applied to a sequence, that sequence is not thereby ungrammatical; it merely lacks the additional enrichment provided by the content of that relation or constraint.[14]

Still, the viewpoint-based coherence relations discussed in this paper are conventions in the same sense that semantic rules of natural

a process need not be naturally divisible into semantic and pragmatic components.

13. There is only this loose parallel between ungrammaticality and "misuse" of coherence relations: when filmmakers clearly signal that a given coherence rule applies, but the logic of the narrative diverges from this rule, viewers may experience a (sometimes brief) sense of incoherence or confusion. But this confusion is an interpretive matter, not a syntactic one.

14. Coherence theorists have standardly claimed that unless *some* coherence relation connects a segment to the rest of the discourse, then we have no connected whole at all, only fragments that happen to be adjacent. Our argument does not require us to commit on this issue, which depends on the sorts of relations one admits.

language are conventions, though the contents of these two classes of conventions may be quite dissimilar. The common notion of convention at work here is illuminated by comparison with the analysis of Lewis (1969, 1975).[15]

Lewis understood conventions as regularities in a population's behavior that are sustained by common practical interest and perpetuated through force of precedent. A standard example is the practice of driving on the right side of the road; drivers share the common interest of avoiding collision, and consistently end up driving on the right (or left) because of their knowledge of prior behavior; it thereby becomes a convention, in Lewis' terms. Crucially, Lewis conjectured that such regularities need not be established through explicit agreement or signaling, but may even emerge from the chance alignment of actions and mental states. Such conventions, like the semantic rules of natural languages, may be acquired and followed without conscious reflection, or even being represented.

In a like manner, we hold that viewpoint constraints can be construed as conventional regularities in the film population's interpretive assumptions. The population here includes both filmmakers and viewers, who, for the sake of reliable communication, have a common interest in coordinating on the interpretive rules (if any) which govern shot-to-shot transitions. There are infinitely many possible rules of coherence to select from, but they only promote communication when a small collection is available to guide interpretation at any one time (Knott and Dale 1994). Given the abiding interest in communication, if one party uses the rules, the other party has reason to do so too; thus, once assumed, they tend to perpetuate, and thereby enter the canon of filmic conventions. To be sure, these are not regularities of action and belief, as Lewis would have it. Rather, they are cognitive regularities: tendencies to associate certain kinds of signals (shot transitions) with certain kinds of contents (viewpoint constraints).

It is finally important to clarify the relationship between conventionality and the concept of "arbitrariness," as we understand it. It is sometimes thought that for a rule to be a convention, particularly a semantic convention, it must be have been selected arbitrarily. But as Metz (1974, 108-10,135-6), Eco (1979, 204-5), and Bordwell (2008, ch. 2) have observed, conventions may in fact be more or less "natural." And some, like the viewpoint constraints canvassed here, are so natural as to be adopted with near universality in global mainstream cinema. We term these NATURAL CONVENTIONS. According to Bordwell, such conventions may be widespread among film cultures, not because we could not rationally do otherwise, but because, given much common psychological and cultural inheritance, certain regularities are natural for beings like us to follow. With little or no explicit learning, these will tend to be the ones that become entrenched as conventions. Indeed, because of this feature, natural conventions may be seamlessly grasped by viewers with no prior exposure, even in the course of interpreting a film.[16]

The same points are reflected in Lewis' analysis, where conventionality merely requires the possibility of an *alternative* convention, but there is no presumption that the choice between a convention and its alternative is arbitrary.[17] In the case of the viewpoint constraints, it is

15. Here we intend to draw on the core insights of Lewis' theory, but we do not endorse the complete theory in every respect. For example, Lewis had specific views about what kind of behavior could constitute a regularity, what it meant for a group to "use" a rule, and what features defined a language (Lewis 1975). Yet these are all vexed issues which we do not wish to take a stand on here. We simply allow ourselves, as a provisional conceptual primitive, the idea that there exist cognitive regularities by which a population comes to use a given rule. And we maintain, with Lewis, that this may arise through the interlocking mental states of participants, even without explicit agreement or endorsement.

16. Armstrong (2016) discusses parallel cases in the linguistic domain in which conventions are learned in the course of interpreting an utterance whose very interpretation relies on them.

17. A given solution may be highly (un)natural but still a "proper coordination equilibrium" in Lewis' sense; so long as, were all participants to settle on that solution, they would be worse off if anyone were to unilaterally adopt an alternative strategy. In the present context, this means that each participant must value reliable communication with other participants *more* than using any particular rule in the course of attempting to communicate.

not difficult to enumerate possible alternatives. For every coherence relation, we can imagine bizarre alternatives which require the opposite kind of coherence to obtain between shots. Filmmakers and viewers must still coordinate, for there are infinitely many possible coherence relations; but what regularity they settle on may— for reasons of psychology, culture, or whatever— be nearly inevitable. This, we believe, is the fate of the viewpoint constraints discussed here: they are natural conventions.

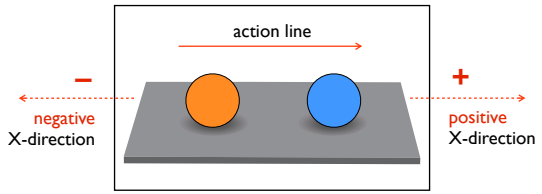
We now turn to the work of this paper: to substantiate the semantic view for the particular case of two viewpoint-based coherence relations: the X-Constraint and the T-Constraint. Since the X-Constraint is already widely discussed, we begin there.

2. The X-Constraint

Our study focusses on the X-Constraint, commonly known as the *180° Rule* or *180° System* (Bordwell et al. 1985; Bordwell and Thompson 2008). Though differing in matters of detail, all formulations share the same kernel idea, that *adjacent shots in a sequence should maintain consistent screen direction for a salient line of action*.[18] The rule is widely disseminated in handbooks on filmmaking, and equally widely followed, at least within mainstream narrative film. Levin and Wang (2009, 35-36) found that 90% of intra-scene sequences randomly sampled from a selection of top-rated mainstream films conform to the X-Constraint.

To define the X-Constraint, we begin with a concept borrowed from film production, that of the ACTION LINE. This is understood as the most prominent linear relationship in a given scene. This may be the path of a particular *action*, like the trajectory of a speeding car, pointed gun, or a glance, but it can also correspond to more static linear arrangements like a wall, road, or, most commonly, a pair of conversationalists.[19] Next we introduce the notion of X-DIRECTION— the direction of the action line, as it is projected along the X-axis of the screen. Independent of its upward/downward or forward/backward orientation, an action line pointing screen-rightward has a *positive* X-direction, while one pointing screen-leftward is *negative*. Since the screen direction of the action line is a function of its apparent relation to a shot's viewpoint (or camera position), we'll say that an action line only has its X-direction *relative* to a given shot's viewpoint.[20] While a given action line may or may not have an inherent direction, it is necessary to arbitrarily assign it one in order to define the concept of *consistency* between action lines, a requirement at the heart of the X-Constraint.

These concepts are illustrated below, where the path connecting the orange and blue balls might be the action line, and we may stipulate that it points in the orange-to-blue direction, and thus has a positive X-direction.



The X-Constraint effectively requires that consecutive shots in a film sequence represent the action line as having consistent X-direction. We will therefore formulate it as a constraint on sequences of two shots

18. We call this the "X-Constraint" because it effectively requires consistent representation of direction along the X-axis of the camera. The term "180° rule" encodes a particular theory of the rule's content, but one we find misleading; more on this in Section 6.

19. Many scenes will contain multiple candidate action lines. (Bordwell and Thompson 2008, 243-4) How one distinguished action line is selected from many possibilities is a pertinent question, and may be resolved by some combination of the filmmaker's intentions, secondary conventions, and common dispositions of the human visual system (Levin and Wang 2009, 37). Resolving this issue is beyond the scope of this essay; for present purposes, we make the simplifying assumption that for every moment in every shot, at most one action line is most salient.

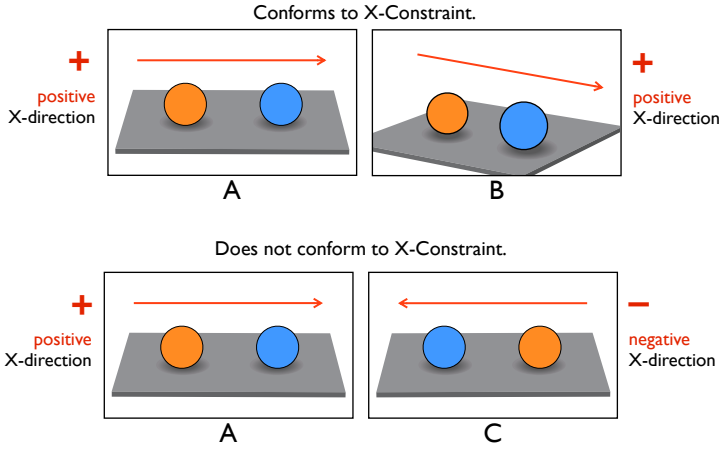
20. A more exact definition of X-direction is furnished in Section 6.

only, though it may be applied iteratively across the course of much longer sequences. For present purposes, we'll state the constraint as follows, leaving a more precise definition for Section 6.

X-Constraint

If the X-Constraint applies to sequence S_1-S_2 then
the X-direction of the action line relative to the viewpoint in S_1
is consistent with its X-direction relative to the viewpoint in S_2.

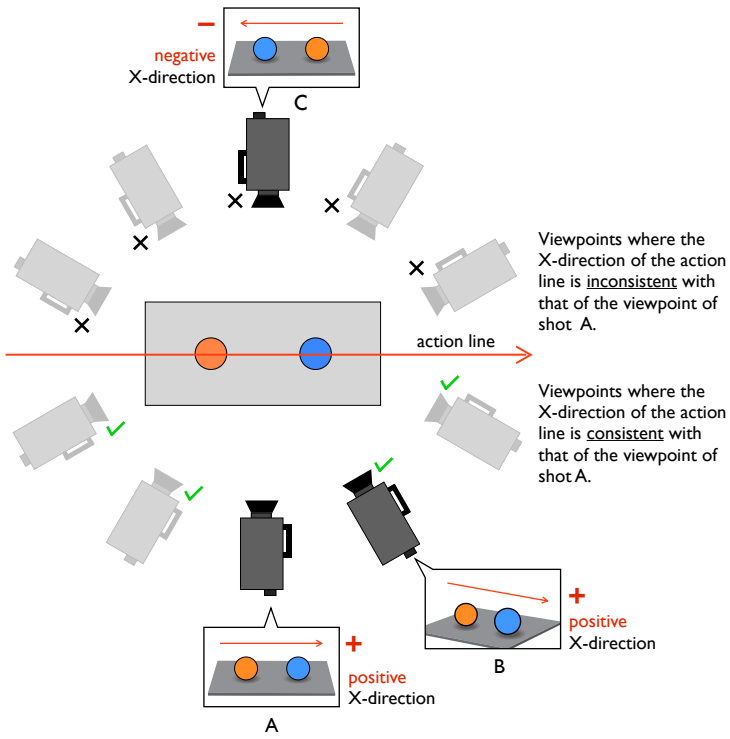
We'll say that an interpretation of a sequence CONFORMS to the X-Constraint just in case it satisfies the consequent of the rule above, whether or not the X-Constraint applies to that sequence. An interpretation FAILS TO CONFORM just in case it does not satisfy the consequent of the rule. By way of illustration, the natural interpretation of the sequence *A*-*B* below conforms to the X-Constraint; for, even though the action line has different screen orientations in each shot, its X-direction remains the same, hence consistent. By contrast, for the sequence *A*-*C*, consider an interpretation according to which the depicted balls do not move between the two shots; then the interpretation fails to conform to the X-Constraint, for although the balls have not moved, the X-direction of the action line is now reversed.[21]



The X-Constraint makes essential reference to "viewpoint"; by this we mean the implied visual perspective or camera position associated with a given shot. We shall understand viewpoints to be oriented locations in space and time, not concrete objects, like eyes or cameras. The same concept of viewpoint applies whether in fact an image was the product of an actual camera, or hand drawn, as in animation. Thus our interest lies strictly in the *implied* viewpoint of a given image, irrespective of how the image was produced. Still, for ease, we will often refer to a viewpoint as a "camera position" and represent it as a camera in illustrative diagrams.

The X-Constraint can be understood as a restriction on the possible position of the viewpoint over the course of a film sequence. To see this, consider the diagram below, illustrating camera positions for the pair of sequences shown above. Besides the *A*, *B*, and *C* cameras, the other shaded out camera positions indicate viewpoints that result in shots which are either X-constraint conforming continuations of *A* (checks), or non-conforming continuations of A (X's).

21. For interpretations where the balls *do* move over the course of the sequence, the action line will also move. In that case, the X-Constraint must apply to the action line at a time t, and the X-direction of the line at t must be the same in both shots (whether or not its direction changes after t, and hence is different at the beginning of the second shot in the sequence). This caveat is dealt with explicitly in the more nuanced definition of the X-Constraint given in Section 6.



As the diagram makes clear, once the position of the initial camera is fixed relative to the action line, the X-Constraint will be satisfied so long as the second camera remains on the same side of the action line. If the second camera is positioned on the opposite side of the action line, the X-direction will have reversed in the resulting image. This gives rise to the filmmaker's heuristic "don't cross the line," as doing so generally ensures conformity to the X-constraint.[22]

Over the course of a film sequence, as shot follows shot, the viewer is presented with one viewpoint and then another. The X-Constraint is what we term a VIEWPOINT CONSTRAINT— a constraint on changes of viewpoint across a cut. It effectively imposes conditions on the evolution of viewpoint as the film progresses, by limiting the possible position and orientation of consecutive viewpoints relative to an action line. These limitations in turn support a more coherent interpretation of the sequence. In practice, filmmakers can rely on viewer's familiarity with the X-Constraint to help convey spatial relations without explicitly depicting them. To follow along, viewers must not only maintain a representation of the depicted world, they must also maintain representations of the viewpoints from which that world is depicted.

Several further notes of clarification are called for at this stage. First, the X-Constraint only comes into play when it is APPLIED; whether it is applied, we will assume, is fact about the film, but one that may depend on the decisions of the filmmaker and the expectations of the viewers. Though it is often applied, it is not typically applied to *every* sequence of shots in a given film. In some shots, there is simply no discernible action line, so the constraint cannot sensibly be applied. Further, the X-Constraint applies only occasionally to transitions at scene breaks; and even within scenes, certain patterns of editing have the effect of suppressing it. In documentary films the constraint is applied much less often than it is in mainstream fiction films, and in some nonfiction genres, such as home movies and news clips, the constraint hardly arises at all. As a rough generalization, within mainstream film, the X-Constraint tends to apply to intra-scene sequences, particularly those where the events represented by each shot stand in specific spatial and causal relations to one another. (But even here there are exceptions, as we'll discuss in Section 5.)

Second, handbooks on filmmaking often construe the X-Constraint as an injunction to filmmakers about what kinds of films to make (or how to make a well-formed film), a stance which emphasizes its kinship with rules of syntax. But this crucially is not our understanding: instead we view it as a rule which *viewers* follow, when they represent it as applied, in the course of film interpretation. In this sense it is a semantic rule: it describes a certain kind of constraint on the mapping from a film to its content, which viewers (who may be filmmakers

22. But this is only an heuristic, as the X-Constraint imposes more specific requirements. See Section 6 for discussion.

themselves) may or may not follow in the course of interpreting the film. It does not imply any direct constraint on what film structures are acceptable, as would a rule of syntax; and it only implies a practical prescription indirectly: filmmakers have an interest in adhering to it insofar as they have a cooperative interest in constructing sequences whose intended interpretation corresponds to the interpretation viewers actually arrive at. If filmmakers neglect it, the risk is not of falling into ungrammaticality, but of causing confusion. As we sill see in Section 5, it is in fact possible to create perfectly intelligible sequences which do not conform to the X-Constraint— a possibility which would be ruled out by a syntactic rule— but filmmakers must take care to signal their intentions appropriately.

Finally, it is tempting to think that the X-Constraint can be identified with perceptual mechanisms that anticipate the motion of the head or eyes. But this cannot be right, though the X-Constraint may, at some level, owe its "naturalness" and ease of use to perceptual cognition. Perceptual rules are characteristically fast and automatic, and relatively unresponsive to world-knowledge, social cognition, or inference. Yet, as we discuss at greater length in Section 5, the application of the X-Constraint can be modulated by scene-breaks, as well as expectations of genre and narrative. Thus it is responsive to properties that are not recognized within the perceptual module, and cannot itself be a perceptual rule.[23]

We hold instead that viewers rely on the X-Constraint as a genuine semantic convention. Yet this position contrasts with what must be the default view: that after the content of individual shots is determined, nearly all representational phenomena in film can be explained in terms of familiar pragmatic processes. Failure to adequately answer this challenge, we believe, has undermined the credibility of many defenses of the semantic view. For example, merely noting that conformity to the X-Constraint is widespread shows little. For this can be explained by the deflationary hypothesis that the constraint is a convention exclusively among filmmakers, like the choice to list credits at the end of the film (Carroll 2008, 119; Cutting 2005, 22). In the remainder of this section, we hope to mount a new and more decisive response to this skeptical challenge.

As for the spatial content that the X-Constraint *seems* to contribute, skeptics hold that this is worked out pragmatically; the rule itself plays no role in interpretation. Currie (1993, 216-7), elaborates the pragmatic view this way, with reference to the alleged convention of shot-reverse-shot editing, a special case of the X-Constraint:

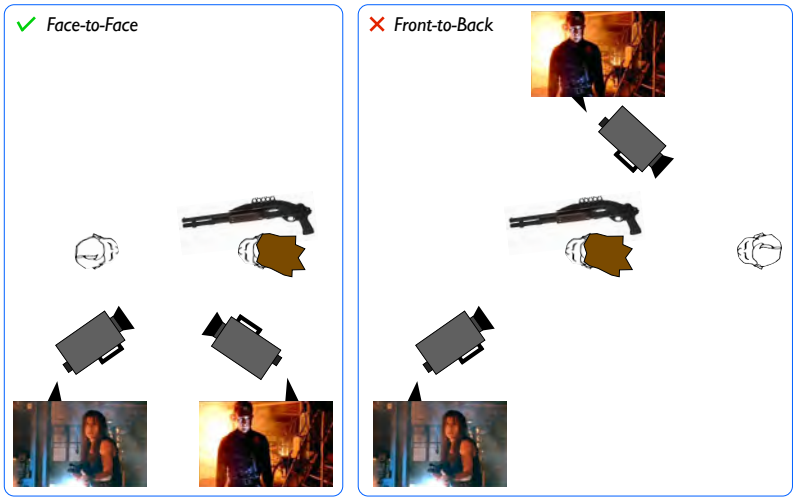
> We infer the connection between the two shots [in a shot-reverse-shot sequence] from the context of surrounding shots, together with assumptions we have made about the course of the story so far, the likely location of the character, and the rationality of the film-maker: we assume that shots and their combinations are chosen by the maker so as to facilitate our comprehension of the story rather than that they succeed one another in an arbitrary fashion. We arrive at a judgement that this is a shot-reverse-shot combination not, as the model of our comprehension of semantic meaning would have it, by understanding a rule of cinematic grammar, but by applying the constitutive rules of rationality...

Thus, on Currie's view, cases of film expression allegedly mediated by the X-Constraint are in fact governed exclusively by "the constitutive rules of rationality."

The plausibility of the pragmatic view is borne out by examples. Consider a shot-reverse-shot sequence from *Terminator 2: Judgment Day* (1991, at 2:21:16), indicated by representative stills:

23. In this respect, it is instructive to contrast the effect of the X-Constraint with the mechanisms which allow viewers to perceive apparent motion in film when exposed to a rapidly flickering sequence of static images. While both mechanisms play a role in film comprehension, only the latter is genuinely perceptual capacity, well-known outside of film scholarship.

Viewers recognize the events depicted in each shot, but also the implied spatial relations *between* these events: that Sarah Connor is shooting *at* T1000 (the android in the second shot), that they are directly facing one another, and so on. Meanwhile, other spatial interpretations are clearly excluded. For example, in the alternative interpretation below, T1000 is coming up behind Connor, and thus they are both facing in the same direction.[24] (The ✓ indicates that *Face-to-Face* is the interpretation viewers naturally come to, and the ✗ indicates that *Front-to-Back* is not an interpretation that viewers naturally come to—not that it is impossible.)



24. Note that the *distance* between the two characters is less precisely established in interpretation than their angular orientation— a fact which is difficult to capture in a diagrammatic idiom.

The preference for the first interpretation over the second *could* be explained by appeal to the X-Constraint. For only in the excluded interpretation, the camera must cross the line of action, thus introducing inconsistent X-directions. But it seems the same conclusion can be derived from our knowledge of the depicted fictional world alone. We know that both Connor and T1000 wish to eliminate one another, that bullets tend to travel in straight lines, and that gun-fights in movies tend to occur face-to-face. It is natural to conclude that the two characters are facing one another. Further evidence comes from continuity in lighting and background elements, and preceding shots in which both characters are simultaneously visible. The pragmatic view holds that all cases like this should be explained likewise, in entirely in pragmatic terms; the X-Constraint plays no explanatory role in interpretation.

What evidence can be attested in favor of the semantic view, and against the pragmatic view? It is not enough to simply *ask* viewers if they are following the X-Constraint, as self-report is a notoriously unreliable guide to actual cognitive processes. More robust evidence has come from psychology, where researchers have repeatedly found that viewers are better able to recall the spatial content of sequences that conform to the X-Constraint than those which do not. (Frith and Robson 1975; Smith et al. 1985; Kraft 1987; Kraft et al. 1991; see Kraft et al. 1991 for a review.) These findings suggest that the X-Constraint plays an important role in the processing and encoding of visual narratives. But such results are inconclusive about the nature of this role— for example, whether the constraint primarily affects the interpretation of film, or the registration of narratives into long-term memory.

We pursue a more direct methodology: we identify a class of cases in which viewers have clear spatial interpretations, yet, unlike the sequence from *Terminator 2* above, this ascription of content can only be explained by appeal to the X-Constraint. But such cases are difficult to find in mainstream film, as filmmakers standardly employ a host of simultaneous clues that overdetermine the intended interpretation of a given sequence. This presumably guarantees faster and more reliable responses from viewers. But for this reason, we must ultimately

appeal to constructed examples to isolate the effects of the conventions we describe.

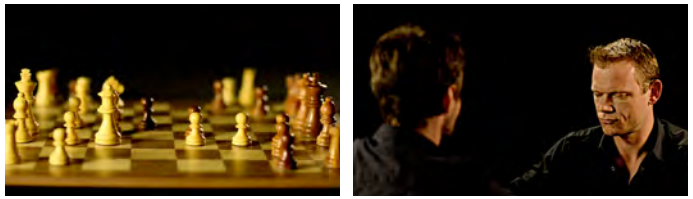
Here we illustrate one such case in detail, which we'll refer to as the "Chess Case."[25] The case is based on a short, author-created film, available here: `http://vimeo.com/73087381`. The film depicts two men playing chess. Stills from each shot are reproduced below. While watching, attempt to answer the prompt: *which side (black or white) is the blond man playing for?*

Viewer judgements decisively favor the answer that, according to the film, the blond player is playing black.[26] Since, on this interpretation, the blond is sitting to the right of the board, relative to the first shot, we will call it *Blond Right*. The alternative interpretation, *Blond Left*, has the blond player playing white.

By design, no single shot favors one interpretation over another, for in no shot is the blond depicted holding a white piece or a black piece, or sitting adjacent to the black side or white side. This is clear in the first two shots, *A* and *B*:

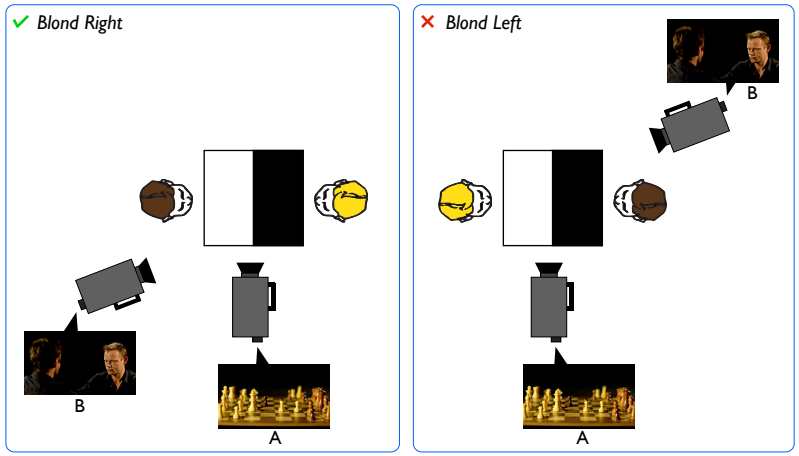
25. The basic structure of the case parallels that of Huff and Schwan (2012), whose findings will be discussed shortly. Whereas Huff and Schwan are concerned to contrast the X-Constraint with a spatial minimization principle, our focus is the contrast between an explanation based on the X-Constraint with a purely pragmatic account.

26. Here we report our own judgements as viewers, and those of many audience members and interlocutors. The experimental findings of Huff and Schwan (2012) support our opinion that these judgements are widespread among mainstream film viewers.



A B

Furthermore, each interpretation is optically consistent, as demonstrated by the positions of cameras and players in the two configurations illustrated below.

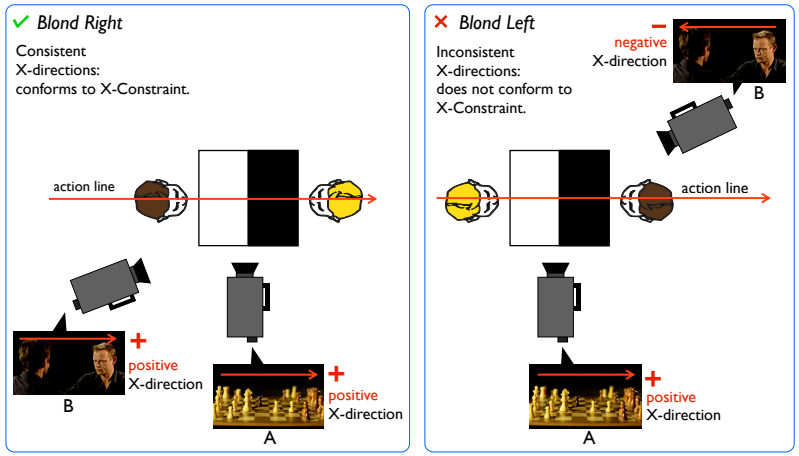


In *Blond Right*, the blond sits on the right side of the board and plays black. The *A*-camera provides a straight-on shot of the board, and the *B*-camera is slightly rotated and displaced to the left. *Blond Left* differs in two ways. First, while the chess board itself stays in the same position, the blond is now seated on the left, playing white. Second, the *B*-camera is now flipped to the opposite side of the board. As a consequence, *Blond Right* and *Blond Left* give rise to precisely the same sequence of camera images.

Despite the fact that each interpretation is consistent with the individual shots, the sequence as a whole clearly communicates the content that the blond plays black. What explains this fact? No purely prag-

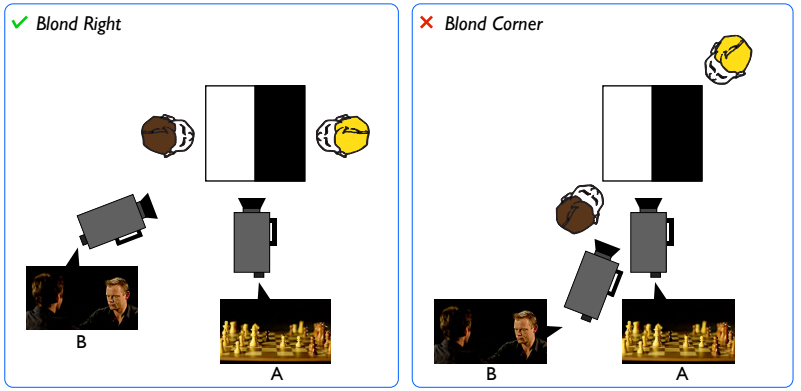
matic explanation seems to be available. There is no narrative context according to which, for example, the blond player strongly prefers to play black. And there is no applicable background knowledge, such as a convention according to which blonds play black and brunets white. Nor, as we have demonstrated, are there any explicit visual clues to draw upon. The pragmatic view, it seems, cannot straightforwardly explain viewers' responses to this case (though we will shortly consider some less straightforward pragmatic proposals).

By contrast, we observe that if viewers are following the X-Constraint, the preferred interpretation of the Chess Case is explained straightaway. To see this, let the salient line of action extend in the positive direction from the left-hand player to the right-hand player (where the direction of the line is arbitrary).



Under the *Blond Right* interpretation, sequence *A*-*B* conforms to the X-Constraint, for the camera does not cross the action line, and as a consequence the X-direction of the action line is positive in both shots, hence consistent. But under *Blond Left*, *A*-*B* does not conform: the camera crosses the action line, with the consequence that the X-direction of the action line is inconsistent between shots. Thus viewers who assume that the X-Constraint applies must rule out *Blond Left*, but not *Blond Right*; this explains the interpretive asymmetry between them.

It should be noted here that while the X-Constraint successfully distinguishes *Blond Right* from *Blond Left*, it is not sufficient to fully *determine* the correct spatial interpretation of the sequence on its own. For example, the Chess Case clearly depicts a scenario in which the blond is sitting on the right side of the chess board, and the brunet on the left side; it does *not* depict a scenario in which each is sitting at one corner of the chess board. This asymmetry is diagramed below. Yet both arrangements are compatible with the X-Constraint, for in neither case does the camera cross the line of action.



To rule out the corner interpretation, we must assume certain facts about how chess is normally played, and about the intentions of the filmmaker. As highlighted in Section 1, this lesson generalizes: the semantic view holds that the conventional rules of film (like the X-Constraint) only determine the content of a film *in concert with* world knowledge, perception, rationality, and other pragmatic resources. In this case, semantics may rule out *Blond Left*, but only pragmatics can rule out *Blond Corner*.

Is there really no way for a defender of the pragmatic view to account for the Chess Case and its ilk? Perhaps pragmatic accounts can avail themselves of richer resources than we have allowed for. Grice

(1975), for instance, developed a set of well-known pragmatic maxims, presumed to be mutually undertaken in any cooperative exchange. Among these, some have invoked the *Maxim of Manner*— an injunction to be as clear, brief, and orderly as possible— in an attempt to derive the standard temporal ordering of narrative discourse (Mauri and van der Auwera 2012, 382-388). They reason that representing the temporal order of events by the temporal order of the sentences that describe them is more orderly than any alternative. And listeners, reasoning in a like manner, will come to expect this.

A Gricean theorist might attempt the same kind of explanation here.[27] Perhaps sequences that preserve screen direction are simply more orderly than others, so are favored for purely pragmatic reasons. The problem with this approach is that there are too many natural types of spatial "orderliness" to choose from. For example, orderly camera motion might involve not moving at all; or moving by a fixed increment in a fixed direction at each shot; or freely translating (but never rotating) from shot to shot; and so on. Each of these relations answers equally well to *a priori* definitions of "orderliness," for each involves a systematic and incremental permutation of spatial features over time. But none has the effect of preserving screen direction in the relevant sense, and none plays a role, as far as we know, in the interpretation of film. (There is one further conception of orderliness, that of a camera which moves *as little as possible* between shots, which might plausibly explain the film data we've reviewed thus far. We address this idea directly below.)

The point here is not that there is no notion of orderliness which could yield the right predictions in this case. Rather, it is that there are too many notions of orderliness, many of which do not yield the right predictions, and rationality alone is insufficient to select among them. Unlike the representation of temporal sequence in narrative, where the one dimensional flow of time naturally suggests a specific notion of discursive orderliness, representations of space are not naturally ordered in a unique way. The only way guarantee the right predictions would be to assume, by fiat, that sequences which conform to the X-Constraint are highly orderly (and that many of the alternative notions considered above are *not*). But this would amount to recognizing the X-Constraint as a stand-alone (natural) convention. Thus the attempt to explain the interpretation as an implicature of manner in this way does not eliminate the appeal to semantic convention.

But there are still other possible explanations of viewer judgements about the film sequences above which diverge from orthodox Griceanism. One set of hypotheses should probably be classed as *semantic* alternatives to the X-Constraint, for they cannot be derived strictly from rational inference and world-knowledge alone. But they are of special interest because they still have broadly pragmatic motivations, and may thus be partially associated with the pragmatic view.

According to the first hypothesis, the displacement of the camera (or implicit viewpoint) from shot to shot must correspond to the physically realistic movement of a human head, eye, or camera from one position to the next. This is the suggestion of Pudovkin (1954, 42): "the lens of the camera replaces the eye of the observer, and the changes of angle of the camera— directed now on one person, now on another, now on one detail, now on another— must be subject to the same conditions as those of the eyes of the observer."[28]

Thus, in the Chess Case, *Blond Left* is ruled out because the film implies no temporal gap between A and B, yet it would be impossible for a physical camera to move from the A-position to the B-position instantaneously.[29] Unfortunately, as Bordwell (2008, 57-60) notes, Pudovkin's conjecture yields unacceptable predictions. In this case, the

27. Thanks to an anonymous reviewer for suggesting this approach.

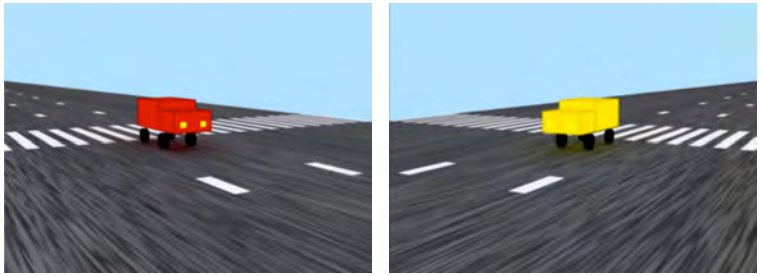
28. This idea is consonant with some versions of Wilson's (2011) "Imagined Seeing Thesis," according to which viewers imagine themselves (including their bodies) to be actually seeing the action of the film; for if this is the case, viewers might expect changes in camera positions to conform to the physical constraints which typically govern human bodies in motion.

29. The film may not require that the A-scene is followed by the B-scene instantaneously. But it is consistent with the content of the film that this is so. The proposed principle would incorrectly rule out this interpretation.

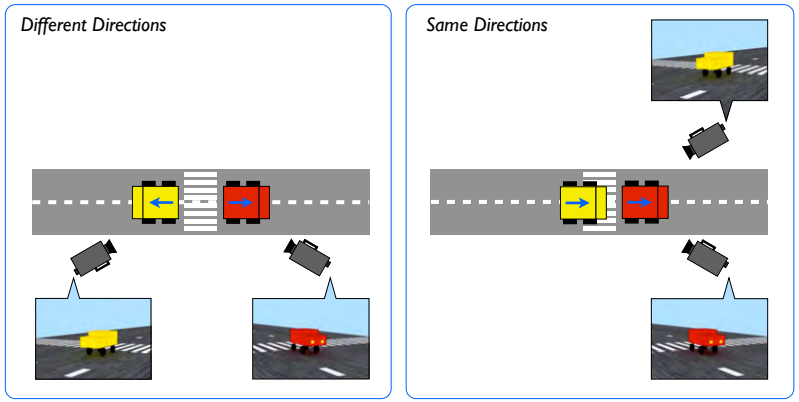
same considerations that bar *Blond Left* would also rule out *Blond Right*, for even though the distance here is smaller, instantaneous displacement remains impossible. (The point could be made more forcefully on a physically larger set!) In general, restricting camera displacement by some kind of normal physical motion simply isn't plausible, since the correct interpretation of many filmic scenes has the camera hopping about from position to position instantaneously. At the same time, we fully recognize that the psychological salience of the X-Constraint may indeed be due, in some indirect way, to in-built perceptual expectations about the motion of the head, or of saccades. Nevertheless, our immediate concern in this paper is with the *content* of the relevant principle, not its psychological origins.

A more serious challenger is a principle we term *Minimize Change*, according to which, if there are several competing pragmatically acceptable interpretations, the one which implies the least change in viewpoint is preferred. The principle might be derived from a domain-general expectation that, all else being equal, objects tend to behave inertially, including the implicit viewpoints of visual sequences. In the Chess Case, for example, it's evident that the distance from camera-*A* to camera-*B* in *Blond Right* is much shorter, in terms of both translational distance and angular rotation, than the corresponding distance in *Blond Left*. Thus Minimize Change correctly predicts the preference for *Blond Right* over *Blond Left*. Yet, we maintain, the principle cannot explain responses in a set of closely related cases, even while the X-Constraint succeeds. Our point is not that Minimize Change has no psychological reality— for it may— but rather that it does not obviate the need for the X-Constraint.

In a recent study, Huff and Schwan (2012) constructed cases which put the X-Constraint into competition with Minimize Change;[30] they found that the X-Constraint better predicts viewer responses than Minimize Change. Audiences were presented with a two-shot animated sequences of red and yellow cars driving on a highway. Here are stills from a representative sequence:



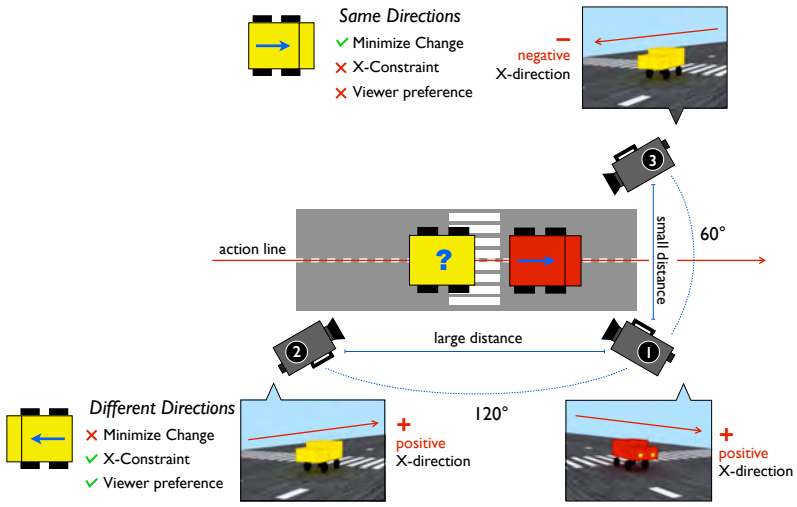
Huff and Schwann asked viewers whether the two cars were driving in the same or different directions, corresponding to the two possible interpretations diagrammed below.



By design, the *Different Directions* interpretation conforms to the X-Constraint, but violates Minimize Change; the *Same Directions* interpretation violates the X-Constraint, but conforms to Minimize Change. The predictions of the two principles, applied simultaneously to both interpretations, are illustrated in the diagram below. Here the 1-camera produces the first shot in both interpretations. The 2-camera produces the second shot in the *Different Directions* interpretation; and

30. Huff and Schwan (2012) call the X-Constraint the "centerline rule" and Minimize Change the "spatial-alignment hypothesis."

the 3-camera produces the second shot in the *Same Directions* interpretation.



As illustrated, the action line lies along the path of the road, so it is clear why *Different Directions* conforms to the X-Constraint, but *Same Directions* does not. On the other hand, Minimize Change favors the *Same Directions* interpretation. To see this, note that the translational distance between 1 and 2 is clearly greater than between 1 and 3 (on the assumption that the same white cross-walk is depicted in each image); in addition, the rotational distance between 1 and 2 is 120°, while that between 1 and 2 is only 60°. Thus, on any reasonable reading of Minimize Change, it recommends the *Same Directions* interpretation.

Huff and Schwan (2012) found that, among viewers in their pool, between 61% and 67% of responses favored the X-Constraint interpretation over the Minimize Change interpretation. Thus, while Minimize Change could in principle explain interpretive preferences in the Chess Case on its own, it cannot do so here, as a majority of viewers take up the interpretation that would be imposed if the X-Constraint applied.[31] In further trials, Huff and Schwann exposed viewers to clips in which the X-Constraint and Minimize Change were *aligned* rather than opposed. Here, between 95% and 99% of responses favored the aligned interpretation. This suggests that Minimize Change (or some similar principle) does in fact influence interpretation, but only alongside, or in concert with, the X-Constraint.

While the preceding argument directly establishes the role of the X-Constraint only in the Chess Case (and Huff and Schwann's car case), it suggests that its use is in fact much more widespread. For if the X-Constraint is generally applied by viewers, at least to intra-scene sequences, it would explain why it is applied here. In sequences like that from *Terminator 2*, whose interpretation does not strictly require the X-Constraint, filmmakers use the convergence of the X-Constraint with pragmatic cues to quickly reinforce the intended spatial meaning.

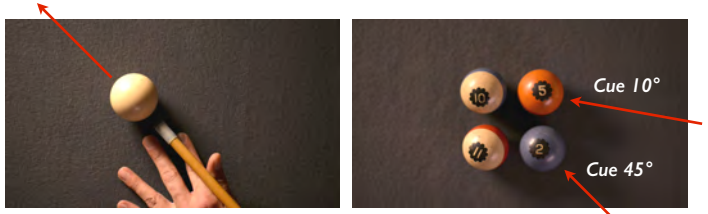
3. The T-Constraint and Beyond

We propose that the X-Constraint belongs to a broader system of semantic rules which constrain the dynamics of viewpoint. In this section, we give some grounds for this conclusion by providing evidence for a novel viewpoint constraint, which we term the Translation Constraint, or T-Constraint for short.

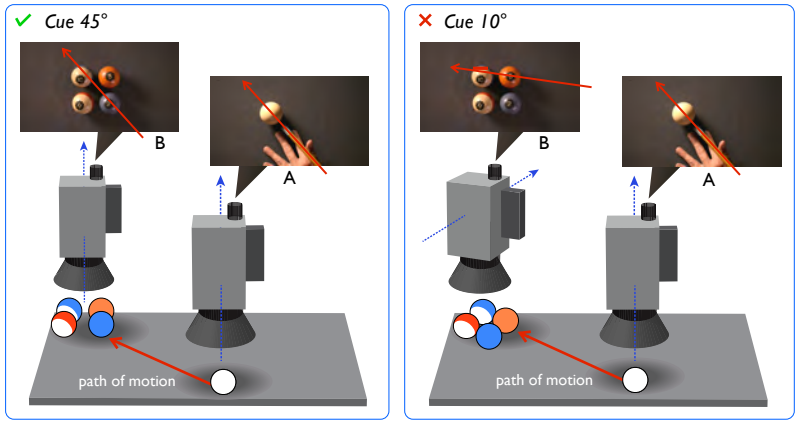
We begin this time with the "Pool Case," based again on a short, author-created film: `http://vimeo.com/51045209`. The sequence depicts a man playing pool. Stills from each shot are reproduced below. While watching, attempt to answer the following prompt: *which pool ball(s) is the white cue ball most likely to strike?*

31. The fact that not all viewers came to an interpretation that conformed to the X-Constraint may be explained by the fact that not all viewers expected it to apply, a point developed further in Section 5.

Viewer judgements recognize as acceptable a narrow band of answers to the prompt question. For example, an acceptable interpretation is that, according to the film, the cue ball would eventually enter the frame at a 45° angle and strike the blue 2 ball. Call this interpretation *Cue 45°*. By contrast, *Cue 10°*, according to which the cue ball would eventually enter the frame at a 10° angle and strike the orange 5 ball, is clearly excluded. The two interpretations are illustrated below:

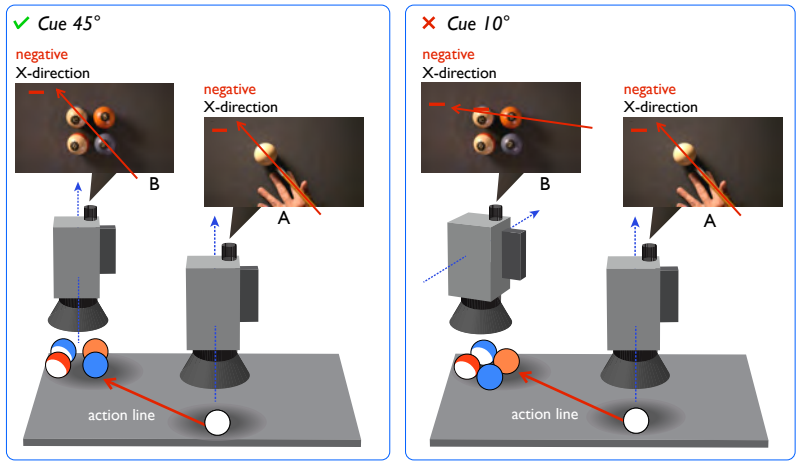


Once again, by design, both interpretations are consistent with the optical and narrative content of the individual shots. This is demonstrated by the positions of the camera and pool balls in the two configurations illustrated below.



In the preferred *Cue 45°*, the *A*-camera provides a shot of the cue ball, with the pool cue positioned at a 45° angle, while the *B*-camera assumes the same orientation over the cluster of pool balls. In the excluded *Cue 10°*, shot *A* is the same, but shot *B* differs in two ways. First, the cluster of target pool balls is now rotated clockwise by 35°. Second, the B-camera is rotated by exactly the same amount in the same direction. The two interpretations yield the same images, but differ with respect to the angle at at which the cue ball would enter shot *B*, and which pool ball it would eventually strike.

As in the Chess Case, we are confronted with an interpretive asymmetry. But in the Pool Case, the X-Constraint cannot explain it, because under both interpretations, the X-direction of the action line— determined by the trajectory of the cue ball— is consistent. (Note, as a corollary, that the camera does not cross the line in either interpretation.)



The X-Constraint fails to adjudicate between *Cue 45°* and *Cue 10°* because, by design, it allows for large variations in angular screen direction from shot to shot, so long as X-direction consistency is maintained. Whatever distinguishes *Cue 45°* and *Cue 10°*, by contrast, must be sensitive to small differences of angular direction.

The T-Constraint does just this: it requires that screen *angle*, not just direction, be maintained in the transition between shots in a sequence. It achieves this by requiring that the second viewpoint in a sequence

be related to the first by translation along a path parallel to the salient line of action, without rotation.[32] We define the constraint as follows:

T-Constraint
If the T-Constraint applies to sequence S_1-S_2 then
the viewpoint in S_1 is related to the viewpoint in S_2
by translation parallel to the action line, without rotation.

The T-Constraint straightforwardly explains what the X-Constraint cannot— why *Cue 45°* is favored over *Cue 10°*. For in *Cue 45°*, the position of the *B*-camera can be derived from the position of the *A*-camera by a translational shift parallel to the action line, without rotation; but in *Cue 10°*, this is impossible, for the camera must rotate.

We conclude that the preferred interpretation of the Pool Case is not merely in *conformity* with the T-Constraint as stated above, but that viewers actually apply the rule to arrive at their preferred interpretation. But what about the Minimize Change principle discussed in the last section? After all, the preferred *Cue 45°* involves less angular change than the dispreferred *Cue 10°*. But Minimize Change cannot account for this preference. On one hand, if Minimize Change requires minimization of both angular *and* translational distance, as it is naturally construed, then it makes false predictions. For then it would require that we interpret the *B*-camera to be located as close as may reasonably be allowed to the *A*-camera. Yet, contrary to this prediction, it is fully compatible with the content of the film that the cluster of four balls is quite close to the cue ball *or* quite far from it.

On the other hand, if Minimize Change is exclusively concerned with angular rotation, then it predicts only that the cue ball will enter the *B*-shot at a 45° angle, but makes no prediction about *where* the ball will enter. Yet viewers prefer an interpretation where the cue will strike the solid 2, and not, say, at the striped 11 or striped 5. A rotation-only variant of Minimize Change cannot account for this response, but the T-Constraint can.

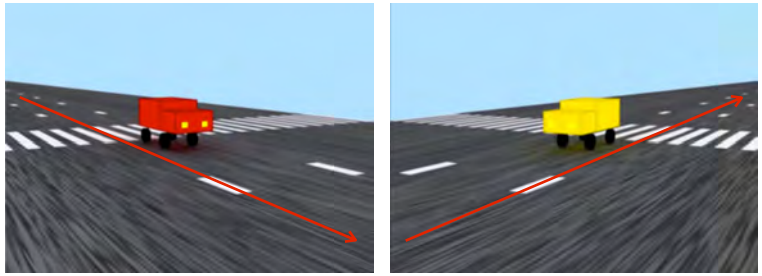
Still the skeptic may not be satisfied; perhaps there is a pragmatic explanation of the same data. One option, again, is to invoke Grice's Maxim of Manner, which prescribes maximizing orderliness, and to maintain that interpretations which are consistent with the T-Constraint turn out to be more *orderly* than alternatives. According to this suggestion, *Cue 45°* is a more orderly interpretation than *Cue 10°* because only the former preserves angular direction of the action line across shots.

But this idea is undermined by the same observations that foil Minimize Change, for it effectively seeks to derive something like Minimize Change from pragmatic considerations. For example, interpretations which conform to Minimize Change would surely be more "orderly" than those which merely conform to the T-Constraint, for they would obey additional spatial constraints in a systematic manner. So a requirement of orderliness would imply minimizing change in many cases. Yet, as we argued above, minimizing change yields the wrong predictions in the Pool Case. Minimizing change is, in effect, *too* orderly; the correct interpretations are often less demanding than that. Thus we conclude that no purely pragmatic explanation accounts for the available data, and the T-Constraint instead plays an essential role.[33]

32. In fact, a variety of other formulations yield the same results in this case, but all pursue more or less the same strategy. For example, an alternative constraint requires that in consecutive shots, the action line have the same position and direction *on the screen*; this permits rotations, where the T-Constraint does not. Adjudicating between these and other closely related proposals requires further data, and is beyond the scope of this paper.

33. An alternative pragmatic hypothesis draws on Sperber and Wilson's (1986) principle of *Relevance*, which requires that contributions to discourse maximize the amount of information which can be extracted from them, while minimizing the cognitive effort required to be interpreted. On this hypothesis, *Cue 45°* is preferred because it provides more information than interpretations which are unspecific about angular relations, and so don't entail, for instance, an answer to the question of which ball is going to get hit (e.g. that favored by the X-Constraint alone), and because it requires less cognitive effort to construct than *Cue 10°* and the like. Our response here mirrors our reply to the Gricean. Interpretations conforming to Minimize Change would be more "relevant" (in the sense of Sperber and Wilson), since assuming minimal change allows one to extract more information from the sequence about the relative proximity

Up to this point, we've argued that at least two distinct viewpoint constraints play a role in determining film content. As they come in and out of application over the course of a film, they work together to help structure its content. Though distinct, the two constraints are logically related. Conformity to the T-Constraint implies conformity to the X-Constraint, since translation without rotation always preserves X-direction. But the reverse is not true, and the example from Huff and Schwann, shown below, illustrates this vividly. In this clip, the viewer-preferred interpretation conforms to the X-Constraint, but clearly does not conform to the T-Constraint, for the action line, which follows the straight path of the road, changes screen-angle sharply from shot to shot. In Section 5 we return to the question of what causes one constraint or another to be applied in different sequences.



While the X-Constraint is familiar to students of film, the T-Constraint, to our knowledge, has never been explicitly recognized. This discovery suggests the existence other viewpoint constraints, defined in terms of the basic spatial permutations that viewpoints may undergo between consecutive shots.[34] And it reinforces the conclusion that tracking the dynamics of viewpoint is a central aspect of film interpretation.

4. Viewpoint Constraints as Coherence Relations

The idea that film is governed by language-like rules has had a long career in film scholarship. Some early theorists took this idea literally, searching for film analogues of syllables, words, and phrases (Pudovkin 1954; Carroll 1980, 72). But these comparisons bore little fruit, leading later thinkers to reject anything like a semantic approach to film (Currie 1993; Carroll 2008, 118-121). Meanwhile, however, Metz (1974) offered a more measured parallel: that between *shots* and *sentences* (not words), and hence that between *sequences* of shots and *sequences* of sentences. He went on to identify film as a species of DISCOURSE, a category that includes any concatenation of meaningful units, such as stories, articles, conversations, or arguments. And he proposed a rudimentary semantic account of film discourse, centered primarily around the structure of temporal relations between shots.

More recently, Bateman and Schmidt (2012) and Wildfeuer (2014) have recast Metz's ideas within a framework borrowed from contemporary formal linguistics. This is the account of linguistic discourse known as DISCOURSE COHERENCE THEORY. (See e.g. Hobbs 1985; Mann and Thompson 1988; Kehler 2002; Asher and Lascarides 2003; Lascarides and Stone 2009b.) Adopting this approach, we argue in this section that the viewpoint constraints which we have discussed should be understood as discourse coherence relations.

The core principles of discourse coherence theory are brought out by an example from Hunter and Abrusán (2016):

(1) I missed my meeting this morning. My car broke down.

In this short narrative, each sentence expresses its own proper content. But the discourse as a whole clearly conveys the additional content that missing the meeting occurred *after* the break down, and that the missing was *caused by* the break down, even though this content is

of the target cluster to the cue ball. So demanding "relevance" would again amount to a demand for minimal change. (Thanks to an anonymous reviewer for suggesting the pragmatic approaches outlined in this note and in the main text.)

34. For example, a "Y-Constraint" might ensure that the direction of gravity maintains a consistent (downward) orientation with respect to the Y-axis of the camera, across shots.

not contained in either component sentence. Discourse coherence theory offers an account of the source of this added content. It proposes that sentences in a discourse are connected together by COHERENCE RELATIONS, informational links that provide the missing content which connects their meanings, and renders the discourse as a whole coherent. Researchers have identified a limited library of coherence relations at work in linguistic discourse.

In the case above, the relationship between the first and second sentence would be classified as a type of *Explanation* relation; this relation requires that the second sentence describe an event which occurs *before* and causes (or explains) the event described by the first.[35] By contrast, the *Result* relation, illustrated below, requires that the second sentence describe a continuation or result of the event described by the first, one that occurs *after* the first:

(2) I missed my meeting this morning. They fired me.

The theory of coherence relations in discourse is an example of a semantic account in an area once thought to be pragmatic (Lepore and Stone 2014, ch. 6). Griceans, for instance, have proposed that the inference of temporal progression exhibited in cases like (2) could be derived from Grice's *Maxim of Manner*: since the sentences are themselves uttered in temporal order, speakers infer that the events so expressed occurred in the same order (Mauri and van der Auwera 2012, 382-388). The story could not end there, of course, because of cases like (1), where the order of temporal interpretation is reversed. Discourse coherence theorists diagnose these as cases of two distinct coherence relations at work. But Griceans might hold that both cases can be accounted for as the result of additional, more complex pragmatic reasoning. However, several lines of evidence ultimately undermine this pragmatic hypothesis.

First, subtle grammatical cues sometimes determine the relation between discourse segments even when that relation is in no sense "rational." In this well-known case from Hobbs (1979, 67), readers invariably infer that the second clause is supposed to explain the first, even though there is no known relevant connection between Istanbul and spinach.

(3) John took the train from Paris to Instanbul. He likes spinach.

According to discourse coherence theory, this is simply another case, like (1), of the *Explanation* relation. As Lepore and Stone (2014, 107) observe, readers "find it extremely compelling to take a follow-up generic stative sentence as an explanation for a prior eventive sentence."

Second, there seems to be a direct connection between particular discourse relations and specific linguistic conventions. Though coherence relations need not be explicitly signaled, they *may* be, with the use of discourse markers like "therefore," "but," "so," "however," and "because." Even more striking, certain words seem to *block* certain relations, in a quite arbitrary manner characteristic of semantic rules. For example, the introduction of the discourse particle "and" into (1) shifts the meaning noticeably, blocking an inference that the second clause explains the first; instead the two clauses are presented as describing merely parallel but independent events:

(4) I missed my meeting this morning. And my car broke down.

By contrast, introducing "and" into (2) has the effect of *reinforcing* the expression of a *Result* relation. It is hard to see how such results could be derived from a purely pragmatic account of interpretation, on which the relevant inference is supposed to survive the basic content being expressed in different ways.

35. The coherence relations of discourse coherence theory are comparable to Metz's *syntagmas*, which govern the temporal (and to some degree, spatial and causal) relationships between shots. And Scott McCloud (McCloud 1993) has proposed a similar framework for understanding the relationship between panels in comics. Lascarides and Stone (2009a,b) extend mainstream discourse coherence theory from purely linguistic discourses to the iconic gestures that accompany speech.

Together, these empirical arguments show that pragmatic inference is neither necessary nor sufficient to determine the relational information conveyed in discourse. Not necessary, because a sentence like (3) can express an *Explanation* relation, without support from world knowledge; and not sufficient, because a sentence like (4) cannot express an *Explanation* relation, even with the support of world knowledge.

We propose that the viewpoint constraints described in this essay, the X-Constraint and the T-Constraint, belong to a family of semantic coherence relations that hold between shots in the domain of film. But whereas previous attempts to apply discourse coherence theory to film have highlighted the temporal relations between shots (Bateman and Schmidt 2012; Wildfeuer 2014), we treat the viewpoint constraints as belonging to a special family of distinctively visual, *spatial* coherence relations. We leave it to future work to settle how these relations are connected to the more familiar rhetorical relations from linguistic discourse.

Understanding viewpoint constraints as coherence relations helps makes sense of certain puzzling aspects of the semantic view of film. Currie (1993), for example, complains that there could not be conventions governing inter-shot relations, because there are no explicit inter-shot markers to be conventionally interpreted. "Compare this with sentential connection displayed in the construction 'P because Q.' In the latter case we can point to conventions of meaning and rules of grammar that determine for this construction a literal meaning: that P occurred because Q occurred." (p. 217)

But this view of the contribution of convention to discourse content can now be seen to be simplistic, restricted as it is to the lexical model of linguistic meaning. By contrast, though discourse relations may be explicitly signaled with a cue word, they need not be. Thus, without explicit marking, (1) expresses an *Explanation* relation, while (2) expresses a *Result* relation. The presence (or absence) of coherence relations is determined by the speaker's intentions, and may be indirectly signaled through a variety of linguistic and contextual cues, such as the shift in grammatical aspect in (3). The case of film is like this. Aside from a few well-known explicit cues (like fading to black), the relations between shots are typically conveyed implicitly, the subject of coordination between filmmakers and viewers.

If our proposal is correct, the comparison between viewpoint constraints and rules of language can now be made precise. Viewpoint constraints are most nearly parallel not to lexical conventions, nor to the compositional rules of subsentential semantics, but instead to those inter-sentential semantic relations which seem to organize all forms of linguistic discourse.

5. Patterns of Application

A familiar challenge for any theory which elevates the X-Constraint to the status of convention is the problem of "violations"— sequences that do not conform to the X-Constraint but do not suffer for it. Carroll (2008, 119-20) cites a well-known example of this phenomenon to support his attack on semantic approaches to film.[36]

> John Ford, for instance, violated the one hundred and eighty degree rule in *Stagecoach* and yet the scene in which this occurs is perfectly intelligible to spectators. No one has ever complained that it was ill formed. They understand that the stagecoach is moving in the same direction after the cut; they do not think it makes a hairpin volte-face. Audiences are able to do this, moreover, because this understanding best coheres with their conception of the rest of the narrative.

An instance of the phenomenon Carroll describes from *Stagecoach* (1939, at 1:13:06) is pictured below. In the first shot, the coach is moving screen-left to screen-right; in the second shot, the screen direction

36. Cutting (2005, 22) draws attention to the same phenomenon. And Currie (1993), in his general criticism of the semantic view, points to analogous "violations" of point-of-view conventions: "there is no convention that says that a shot of a character face-on followed by a different shot means that the second shot is from the point of view of the character — there are too many cases where a shot of the first kind is followed by a shot that is not subjective."

reverses, but with no loss of spatial perspicuity. As Carroll attests, such examples may be found throughout the canon of mainstream cinema, up through contemporary film.

We understand Carroll's objection this way: if the X-Constraint is a semantic convention, then it is expected to apply in intra-scene sequences; and if it applies, then non-conforming interpretations should be ruled out. Yet there are cases of intra-scene sequences whose preferred interpretations clearly do not conform to the X-Constraint: in the case at hand, viewers "understand that the stagecoach is moving in the same direction after the cut." Thus, Carroll concludes, the X-Constraint cannot be a semantic convention.[37]

Our initial reply to the argument is simple: as a matter of fact, the X-Constraint does *not* apply to the sequences in question, hence there is no presumption of directional continuity, and no threat of counterexample. These are not "violations" after all. Whether a given constraint applies or not is largely a matter of the filmmaker's intentions, and these are cases where the filmmaker clearly intends the rule not to apply.

37. Some of Carroll's remarks suggest an additional objection. This one begins with the assumption that if the X-Constraint were real, it would take the form of a *syntactic* rule, a constraint on how films may be properly constructed. But since films which fail to conform to the rule are not, as Carroll observes, judged to be "ill-formed," there can be no such rule. In response, we remind readers that we view the X-Constraint as a *semantic* rule, and not a syntactic one. Films may be freely put together without the X-Constraint (or any other coherence relation) with no interruption of grammaticality.

The principle that a filmmaker's intentions generally determine which convention is applied is inherited from the philosophy of language. For although a speaker's intentions cannot alter the substance of a given linguistic convention (for example, what a word can mean), they *do* as a first approximation determine which convention is in effect. Such intentions are thought to fix the meaning of ambiguous terms or the reference of deictic pronouns (Kaplan 1989). In this respect, film is no different: it is not up to the filmmaker what the X-Constraint *is*, but it is largely up to the filmmaker to decide whether the X-Constraint *applies*, or not.

As for viewers, the correct spatial interpretation in such cases is secured without help from the X-Constraint, through pragmatic inference alone: we know the coach in *Stagecoach* is fleeing from its attackers, and that the only way to do this is to continue in a straight line across the plain. The natural inference is that both shots depict the same worldly direction of motion, from opposing camera angles. The filmmaker, knowing we would draw this inference, may intentionally eschew the X-Constraint, without fear of inducing spatial confusion.

In itself, the claim that there are sequences which are not in the domain of the X-Constraint is unproblematic. As we've observed, the X-Constraint is often absent in the transition from one scene to another, and may be freely ignored in genres outside of mainstream film. More generally, there is no problem with the notion of a semantic convention that applies in certain cases and not in others. For instance, the linguistic signal 'bank' denotes a financial institution according to one convention, and the edge of a body of water according to another. On some occasions one semantic convention is engaged, and sometimes the other, and there is nothing unusual about this selective deployment. In fact, such optionality is the norm for the specific case of coherence relations, which do not apply uniformly to every segment of a discourse. Indeed, in Section 4, we saw one discourse where the *Explanation* relation (but not the *Result* relation) applied, and another displaying the opposite distribution. In general, so long as interlocutors have *some*

mechanism for coordinating on which convention applies to a given segment of discourse, selective application is unproblematic.

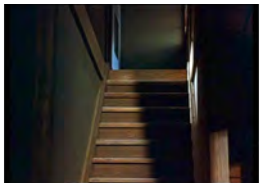
But if the X-Constraint doesn't apply in cases like that from *Stagecoach*, how do viewers come to know this, especially given that these are intra-scene sequences where the constraint normally *does* apply? The answer, we think, depends in part on the fact that viewers typically expect filmmakers to tell more or less coherent and logical stories. Assuming the X-Constraint in the *Stagecoach* sequence would lead to absurd narrative content, implying that the coach had nonsensically reversed direction with impossible rapidity. Faced with the choice between imputing absurd narrative intentions to the filmmaker, or merely ascribing the intention not to apply the X-Constraint, viewers opt for the latter. They do so even if the context of the film is one where the positive application of the convention would normally be expected. Thus, genre and scene structure permitting, viewers seem to assume as a default that the X-Constraint is applied, unless doing so leads to absurdity. Furthermore, filmmakers realize this and design their films so that the inference to non-application, if it is to be made, is easy and clear. This account would explain the instances of positive application we have seen, as in the Chess Case, and non-application, as in the example above.

In certain cases of unsuccessful editing, failure to apply the X-Constraint gives rise to a sense of confusion; such sequences don't "cut together." Unlike the example from *Stagecoach*, we submit, these are cases where the filmmaker has specifically led the viewer to *expect* the X-Constraint to apply, but it does not. The process of having one's spatial expectations violated gives rise to the characteristic feeling of confusion, and the disruption may lead to disorientation with respect to the space depicted by the film.

The same basic account of selection application carries over to interactions between the X-Constraint and T-Constraint. Recall the sequence from Huff and Schwann, where adjacent shots depict different cars driving on the same road, but from vantage points separated by some degree of rotation. As we saw, in the preferred interpretation of that example, the X-Constraint, but not the T-Constraint applies. As predicted, assuming that the T-Constraint applies in that case would lead to a bizarre, spatially disjointed content for the whole (involving a chaotically moving roadway). Viewers expect filmmakers to depict spatially coherent situations, so conclude that the T-Constraint does not apply.

While the discussion thus far suggests an elegant picture of constraint application, it is ultimately too simplistic. For in a prominent class of cases, the X-Constraint is not applied, though its application would not lead to any incoherence. These are what McCloud (1993) has called "aspect-to-aspect" sequences. Such sequences are used to reveal static facets of some unified physical setting, but without pushing the central narrative forward. Aspect-to-aspect editing gives rise to content with "low" spatial coherence: the scenes depicted make spatial sense, but do not imply any specific spatial relationship between shots, save that they all depict the same general locale.[38]

An example from Yasujirô Ozu's *An Autumn Afternoon* (1962, at 1:53:11) illustrates such aspect-to-aspect editing. Drawn from the final scene of the film, the sequence reveals the protagonist's empty house after his daughter marries and moves out. We are presented with a series of shots, each of which depicts some aspect of the house, yet they are not linked by any more specific spatial relations than this. The sequence doesn't obviously preclude conformity to the X-Constraint, but it is not part of our preferred interpretation that it conform to it either.


38. See Metz (1974, 127-8) for similar ideas; Metz classifies such editing as an example of the "descriptive syntagma."

Such examples differ from explicitly non-conforming cases, like the sequence from *Stage Coach*. For in aspect-to-aspect editing, the X-Constraint is not assumed, but assuming it would lead to no absurd or surprising interpretation. One plausible suggestion holds that viewers do not assume the X-Constraint here because applying it would do nothing to further the narrative coherence of the sequence. An alternative idea is that the X-Constraint is applied by default, not in intra-scene sequences generally, but in a more specific species of sequences which exclude aspect-to-aspect editing. It is noteworthy that that in typical aspect-to-aspect editing, no identifiable action line overlaps consecutive shots, and there is often no causal continuity across the sequence. This marks a clear contrast with standard sequences in which the X-Constraint applies, which typically include both action lines and obvious causal continuity. So viewers may be drawing on a variety of clues to reach the conclusion that the X-Constraint does not apply in a given instance.

If any one of these hypotheses is right, the X-Constraint would appear not so much a simple default assumption as an inference licensed by subtle signaling on the part of the filmmaker. We do not attempt to resolve the issue, but note that the problem we face here is quite general. Even within discourse coherence theory in linguistics, there is little consensus about the mechanisms which determine patterns of application for different coherence relations. Hobbs et al. (1993) and Asher and Lascarides (2003) have offered alternative accounts of the non-monotonic reasoning which grounds the application of discourse relations, but this remains an area of open and fruitful research.

In sum, we have argued that exceptions to the X-Constraint and T-Constraint are not counterexamples, but merely instances of non-application. And we have indicated some of the factors that make application or non-application more likely in particular cases. Though we have not provided a full theory of application for viewpoint constraints, we hope that we have said is sufficient to defend the X-Constraint against persistent objections.

6. The Semantics of Viewpoint Constraints

In this final section we provide a more detailed semantic analysis of the viewpoint constraints discussed above, with an eye toward formalization. After expanding the key definitions, we compare the present analysis to the standard "'180°'" or "Don't Cross the Line!" formulations, and conclude with a sketch for implementing our semantics within a possible-worlds framework. Our focus is on the X-Constraint, but the same kind of formalization may be applied to the T-Constraint.

We begin by recalling the original definition of the X-Constraint given in Section 2:

X-Constraint (preliminary)
If the X-Constraint applies to sequence S_1-S_2 then
the X-direction of the action line relative to the viewpoint in S_1
is consistent with its X-direction relative to the viewpoint in S_2.

The definition glosses over two important facts related to motion. First, within each shot, the viewpoint may continuously shift position, as is typical of a moving camera. This is accounted for by allowing each shot to be associated with a continuum of viewpoints over time. Problems of spatial coordination only arise at the transition *between* shots. The X-Constraint helps to solve this problem by acting as a link between the last viewpoint of the first shot, S_1, and the first viewpoint of the second shot, S_2, for these are the viewpoints which bookend the transition between shots, and the only ones in the sequence which require some extra-perceptual mechanism to be coherently connected.

The second issue is that, within each shot— or between them— the action line may itself shift position, independent of the motion of the camera, as when the central character in a dinner scene shifts her gaze from a person on one side of the table to a person on the other. This is resolved by specifying that it is only the position of the action line as depicted by the last moment of S_1 which is relevant. Viewers track the position of the action line throughout a shot, up to its final moment. After the cut, the last position and angle of the action line serves as

a kind of spatial anchor to ground the interpretive leap into the next shot. The key assumption here is that the action line depicted in one shot, considered as an abstract geometrical form, enters into the space depicted by the subsequent shot— even if the objects that originally defined the line have moved or disappeared. This allows viewers to compare the direction of the same, fixed action line in both shots.

With these amendments in mind, we can articulate a final definition:

X-Constraint (final)
If the X-Constraint applies to sequence S_1-S_2 then
the X-direction of *a* relative to the *last* viewpoint in S_1
is consistent with
the X-direction of *a* relative to the *first* viewpoint in S_2,
where *a* is the position of the action line last depicted in S_1.

To define X-direction, we must understand each viewpoint as establishing a three-dimensional coordinate frame, with the Z-axis running forwards and backwards through the camera, the Y-axis running up and down, and the X-axis running left and right. The X-direction of an action-line relative to a viewpoint is defined as the sign (positive or negative) of the X-coordinate of the unit vector pointing in the direction of the action line.[39] Since action-line and viewpoint are both embedded in the same depicted space of the film, this comparison will always be possible.

We recognize three X-directions: *positive*, *negative*, and *null*. Two such directions are CONSISTENT if they are not opposites.[40] As a consequence, a null direction is consistent with any other direction. This captures the idea that "on-the-line" shots may be cut together with a viewpoint on either side of the action line. This is because a viewpoint whose front-to-back (or Z-) axis aligns with the action line has a null X-direction, which is consistent with either a positive or negative X-direction. As Bordwell and Thompson (2008, 242-3) note, this fact helps explain why directors often use intermediate on-the-line shots as a technique that allows the camera to ultimately cross the action line while maintaining conformity to the X-Constraint on a cut-by-cut basis.

The heart of our definition of the X-Constraint— the idea of restricting camera motion relative to an action line— derives from the traditional "Don't Cross the Line!" formulation of the 180° Rule. But our characterization of *consistency* in particular diverges from it. Roughly, the 180° Rule requires that consecutive viewpoints occupy positions within the same 180° hemisphere defined by the action line (as illustrated by the figure on page 9). While this idea has proven adequate for many purposes, we have found that our approach — defining acceptable transitions in terms of consistent viewpoint directions— has several advantages. The primary virtue of the X-direction formulation is perspicuity. It allows for an elegant and modular account of the X-Constraint, where other constraints can be easily derived from or compared to it, by imposing additional restrictions on shifts along the Y- and Z-axes, for example, or additional constraints on translation, as with the T-Constraint. Furthermore the X-direction formulation more directly reflects the idea of maintaining consistent screen direction than the 180° formulation.

A second advantage is its handling of certain liminal cases. Consider a sequence where the camera in shot *A* is positioned directly over the action line, as in the first shot of the pool case. Shot *B* is also positioned directly over the action line (perhaps at a different point along the line), but is rotated 180°, so that the action line now has the opposite screen direction. According to the 180° Rule, since both cameras are above the line, hence in the same 180° hemisphere carved out by it, the rule is satisfied. Nevertheless, we submit, such cases, which involve reversals of X-direction, intuitively violate a requirement for consistent screen direction which the X-Constraint captures. (Similar cases are generated when the camera faces the action-line from the side.

39. This coordinate may be calculated as the dot product of the vector in the direction of the action line and the vector in the direction of the X-axis.
40. Mathematically, two directions are consistent if the product of the dot products (i.e. X-coordinates) from the two shots is nonnegative.

The 180° Rule allows the camera to rotate fully upside down, but the X-Constraint prohibits this.) Thus it appears to us that the 180° Rule is a practicable heuristic for filmmakers seeking to preserve consistent screen direction, but it is merely an approximation of the X-Constraint which is more precise, general, and fundamental.

We conclude by providing a sketch for how the current proposal may be situated within the framework of a possible worlds semantics. As a first step, we think of the content of a single static image as a unified PICTURE SPACE, populated with an arrangement of individuals, properties, and relations, at a single moment in time (Howell 1974). This space is centered on a viewpoint, which is understood as specifying both a location in space and time, and an orientation in the depicted space. The objects and properties depicted by the image are located in the picture space relative to this viewpoint (Hopkins 1998, ch. 3; Hyman 2006, ch. 5). Such a space may in turn be modeled as a set of $\langle world, viewpoint \rangle$ pairs, or "viewpoint-centered worlds" (Ross 1997, 73; Blumson 2009). Each world must be compatible with the space described by the picture, relative to that world's associated viewpoint.

Whereas the content of a static image is a particular viewpoint-centered space at a single moment in time, a dynamic image, or shot, depicts a changing space over an interval of time. We may then understand the content of a shot using the notion of a VIEWPATH— a continuous sequence of viewpoints, one for each moment depicted by the shot. The content of a shot is then modeled as a set of $\langle world, viewpath \rangle$ pairs. To determine the content of a shot is to determine what kind of world it depicts, and to determine what kind of path the viewpoint took through that world. As with a static image, the shot will never have as its content a maximally specific single world and viewpath, but instead describes a range of possible worlds and viewpaths through them, corresponding to the way it represents some but not all aspects of the world depicted.

Finally, where an individual shot depicts the world from a single, continuous viewpath, a film *sequence* depicts the world from *series* of disconnected viewpaths, which will typically have different temporal and spatial locations within that world. And the content of a film sequence as a whole can be modeled as a set of pairs of worlds and *sequences* of viewpaths.

Following a tradition in formal semantics, we understand sequential interpretation as progressive elimination of possibilities. In principle, a given sequence is compatible with any number of interpretations, where each "interpretation" corresponds to a unique world and sequence of viewpaths. To correctly interpret a sequence is to eliminate all those possible interpretations which are incompatible with its content. This understanding is prefigured in our discussions of the *Terminator*, Chess, and Pool cases, where interpretation is cast as a process of eliminating world-viewpoint assignments which are not compatible with the sequence's content.

The most basic kind of constraint on a sequence's content comes from the contents of the composing shots themselves. The content of the sequence includes both the spaces depicted by each component image as well as the series of viewpoints from which these spaces are depicted. With some idealization, this assumption can be formalized as follows, for any two-shot sequence S_1-S_2. (Here we use $[\![\phi]\!]$ to denote the content of ϕ.)

(5) For any $\langle w, \langle v_1, v_2 \rangle \rangle \in [\![S_1\text{-}S_2]\!]$:
$\langle w, v_1 \rangle \in [\![S_1]\!]$ and $\langle w, v_2 \rangle \in [\![S_2]\!]$.[41]

That is, for any world w and viewpath sequence $\langle v_1, v_2 \rangle$ in the content of S_1-S_2, w and v_1 must be in the content of S_1, and w and v_2 must be in the content of S_2.

With these components in place, it is easy to construe the X-Constraint as simply one more assumption which, like (5) limits the

41. A more general version of the constraint may be stated recursively. For any sequence of shots S_s and shot S_n: for any $\langle w, \langle v_1, ...v_n \rangle \rangle \in [S_s\text{-}S_n]$: $\langle w, \langle v_1, ...v_{n-1} \rangle \rangle \in [S_s]$ & $\langle w, v_n \rangle \in [S_n]$. The same approach may be extended to the X-Constraint.

content of an arbitrary sequence. Under this construal, we may formulate the X-Constraint as follows:

(6) If the X-Constraint applies to sequence S_1-S_2 then
for any $\langle w, \langle v_1, v_2 \rangle\rangle \in [\![S_1\text{-}S_2]\!]$:
the X-direction of *a* in *w* relative to the *last* viewpoint in v_1
is consistent with
the X-direction of *a* in *w* relative to the *first* viewpoint in v_2
where *a* is the position of the action line at *w*, from v_1,
at the last moment of v_1.

Put another way: $[\![S_1\text{-}S_2]\!]$ is a subset of the set of pairs of worlds and viewpath sequences where each of those pairs corresponds to an interpretation which conforms to the X-Constraint. Of course, in a full formalization— say, in a computer model— much more would need to be said about the structure of the possible worlds in question, the identification of the action line, and so on. But we hope to have said enough to indicate how one might proceed here.

7. Conclusion

We have argued for a version of the semantic view of film representation, and we have put forward the X-Constraint and T-Constraint as substantial hypotheses of this approach. Our defense of the claim that viewpoint constraints are semantic conventions in film has taken different tacks. Through the use of the Chess Case and the Pool Case, we argued that theorists are compelled to recognize the X-Constraint and T-Constraint in order to explain regularities of interpretation among viewers. In addition, we have argued that such explanations are at least *viable*, by defusing the threat of counterexample, and even *plausible*, by contextualizing them within the broader framework of coherence relations, and by showing that they may be formalized in a manner that predicts the judgments in our cases. If film is to have a language, we conclude, it is a language made up of rules like these.

References

Armstrong, J. (2016). The problem of lexical innovation. *Linguistics and Philosophy*, 39(2):87–118.

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

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

Blumson, B. (2009). Pictures, perspective, and possibility. *Philosophical Studies*, 149(2):135–151.

Bordwell, D. (2008). *Poetics of Cinema*. Routledge New York.

Bordwell, D., Staiger, J., and Thompson, K. (1985). *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. Columbia University Press.

Bordwell, D. and Thompson, K. (2008). *Film Art, 8th Edition*. McGraw-Hill, New York, NY.

Carroll, J. M. (1980). *Toward a Structural Psychology of Cinema*, volume 55. Walter de Gruyter.

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

Currie, G. (1993). The long goodbye: The imaginary language of film. *British Journal of Aesthetics*, 33(3):207–219.

Currie, G. (1995). *Image and Mind: Film, Philosophy and Cognitive Science*. Cambridge University Press.

Cutting, J. (2005). Perceiving scenes in film and in the world. In Anderson, J. D. and Anderson, B. F., editors, *Moving Image Theory: Ecological Considerations*. Southern Illinois University Press.

Eco, U. (1979). *A Theory of Semiotics*. Indiana University Press, Bloomington.

Frith, U. and Robson, J. (1975). Perceiving the language of films. *Perception*, 4(1):97–103.

Grice, P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics, Volume 3: Speech Acts*, pages 41–58. Academic Press, New York.

Harman, G. (1977). Semiotics and the cinema: Metz and Wollen. *Quarterly Review of Film & Video*, 2(1):15–24.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science*, 3(1):67–90.

Hobbs, J. (1985). *On the Coherence and Structure of Discourse*. CSLI.

Hobbs, J., Stickel, M., Martin, P., and Edwards, D. (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Hopkins, R. (1998). *Picture, Image and Experience*. Cambridge University Press, Cambridge.

Howell, R. (1974). The logical structure of pictorial representation. *Theoria*, 40(2):76–109.

Huff, M. and Schwan, S. (2012). Do not cross the line: Heuristic spatial updating in dynamic scenes. *Psychonomic Bulletin & Review*, 19(6):1065–1072.

Hunter, J. and Abrusán, M. (2016). Rhetorical structure and quds. *Unpublished.*

Hyman, J. (2006). *The Objective Eye: Color, Form, and Reality in the Theory of Art*. University of Chicago Press, Chicago.

Kaplan, D. (1989). On demonstratives. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Center for the Study of Language and Information (US).

Knott, A. and Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35–62.

Kraft, R. (1987). Rules and strategies of visual narratives. *Perceptual and Motor Skills*, 64(1):3–14.

Kraft, R., Cantor, P., and Gottdiener, C. (1991). The coherence of visual narratives. *Communication Research*, 18(5):601–616.

Lascarides, A. and Stone, M. (2009a). Discourse coherence and gesture interpretation. *Gesture*, 9(2):147–180.

Lascarides, A. and Stone, M. (2009b). A formal semantic analysis of gesture. *Journal of Semantics*, 26(4):393.

Lepore, E. and Stone, M. (2014). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.

Levin, D. and Wang, C. (2009). Spatial representation in cognitive science and film. *Projections*, 3(1):24–52.

Lewis, D. (1969). *Convention: A Philosophical Study*. Blackwell, Oxford.

Lewis, D. (1975). Languages and language. In Gunderson, K., editor, *Language, Mind and Knowledge*, volume 7 of *Minnesota Studies in the Philosophy of Science*, pages 3–35. University of Minnesota Press, Minneapolis.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *Text*, 8(3):243–281.

Mauri, C. and van der Auwera, J. (2012). Connectives. In Allan, K. and Jaszczolt, K., editors, *The Cambridge Handbook of Pragmatics*. Cambridge University Press.

McCloud, S. (1993). *Understanding Comics: The Invisible Art*. Kitchen Sink Press.

Metz, C. (1974). *Film Language: A Semiotics of the Cinema*. University of Chicago Press.

Nöth, W. (1995). *Handbook of Semiotics*. Indiana University Press.

Pudovkin, V. I. (1954). *Film Technique and Film Acting*. Vision Press Limited, London.

Ross, J. (1997). *The Semantics of Media*. Kluwer Academic Publishers.

Smith, R., Anderson, D., and Fischer, C. (1985). Young children's comprehension of montage. *Child Development*, pages 962–971.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and Cognition*. Harvard University Press, Cambridge, MA.

Stalnaker, R. (1978). Assertion. In Cole, P. and Morgan, J., editors, *Syntax and Semantics, Volume 9: Pragmatics*, pages 315–32. Academic Press, New York.

Wildfeuer, J. (2014). *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*, volume 9. Routledge.

Wilson, G. (1989). *Narration in Light: Studies in Cinematic Point of View*. Johns Hopkins University Press.

Wilson, G. M. (2011). *Seeing Fictions in Film: The Epistemology of Movies*. Oxford University Press.