An existential semantics for over-informative attitudinal embedding in film[1]

Mats ROOTH — *Cornell University*
Dorit ABUSCH — *Cornell University*

Abstract. We review prior work on geometric semantics, temporal progression, and indexing in pictorial narratives, and then develop them for film, using an abstract syntax that assembles film shots, and is interpreted compositionally. The result is a compositional possible worlds semantics for extensional passages in film. Intensional passages in film include shots describing the hallucinations, dreams, and recollections of a character. Thematically, they include stories about drug-induced hallucination, characters who see the dead, and schizophrenic characters who interact with imaginary ones. In the abstract syntax, a straightforward embedding strategy is adopted, involving an attitude predicate, an index for an experiencer, and an index for the counterpart of the experiencer in the embedded passage. For the semantics, we argue that there is a systematic problem of the embedded shots having detailed geometric, temporal, and (for sound film) acoustic information. This information is so strong that it is implausible that in described situations, the experiencing character should have strong enough information to entail the information in the embedded shot. This makes Hintikka/Lewis semantics for embedding in film problematic. A starting point for a solution is found in linguistic descriptions of the content of embedded film passages that use an existential modal *could*, rather than a universal one. This is developed into a possible worlds semantics for filmic embedding that uses existential world quantification, strengthened by a normality condition stated using an ordering source.

Keywords: Daydreaming, de se, film, geometric semantics for film, embedding, hallucination, Hintikka semantics for attitudes, Lewis semantics for attitudes, possible worlds semantics, premise semantics for modality.

1. Introduction

In film and series, passages that are plausibly analyzed as embedded are common. In the streaming series *Badehotellet*, the character Vera in a séance sees and converses with a dead man, Mr. Andersen (Figure 1). In the film *Eileen*, the protagonist, who works in a prison, has a propensity to daydream. In an early scene, she daydreams about a guard Randy striding over and making out with her (Figure 2). In the film *A Beautiful Mind*, the protagonist Nash interacts with a grad school roommate Charles, the roommate's niece, and a defense department official. They exist in his imagination, not in a world consistent with the film (Figure 3). These daydreams and hallucinations are described by simple film shots, which are not distinguished in any direct way from extensional shots. We take the straightforward approach of positing embedding for these shots. In the framework we assume, these shots cannot be unembedded, because then worlds described by the films would have to look like the shots.

We assume a setup where a film shot contributes a model-theoretic object constructed from worlds and individuals. When a shot is embedded, constraints on base worlds of the film should

Figure 1: In a séance Vera Madsen sees and converses with a man whose death she may be responsible for. *Badehotellet*, S2 E4 (2015). Vera Madsen (Alberte Blichfeldt), Lydia Ploug (Sonja Oppenhagen), Julius Andersen (Ole Thestrup).
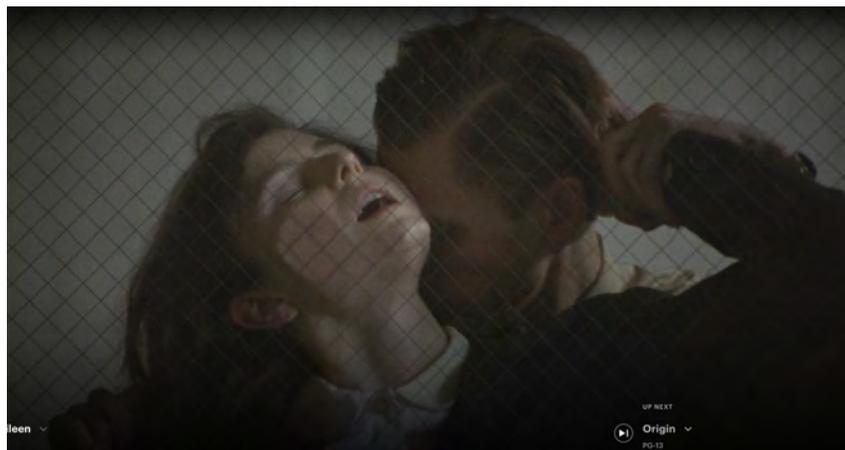

Figure 2: Eileen daydreams about making out with Randy. *Eileen* (2023). Eileen (Thomasin McKenzie), Randy (Owen Teague).

Figure 3: Nash interacts with a hallucinated grad school roommate Charles. *A Beautiful Mind* (2001). Charles Herman (Paul Bettany) and John Nash (Russell Crowe).

somehow be imposed referring to the content of the embedded shot. In natural language semantics for embedded sentences, this is standardly done with entailment semantics, where the content of the attitudinal state of the agent in a described world is required to entail the content of the embedded sentence. We argue that entailment semantics for embedding is unworkable for film, because the geometric and acoustic content of film shots is very strong. It is implausible that the information of an agent in any plausible scenario could be strong enough to entail this content. In response, an initially weaker semantics is proposed, where rather than entailing the embedded information, the information of the agent merely has to be compatible with it. By itself this is unacceptably weak on its own, but it can be strengthened by a normality condition.

The paper is organized like this. Section 2 outlines a framework for the possible worlds semantics of film. Section 3 introduces an event semantics for attitudes such as hallucination, gives empirical arguments against using a Hintikka/Lewis semantics for attitudinal embedding in film, and starts to suggest an alternative with weakened quantificational force for embedding. Section 4 presents our analysis, which uses an existential semantics for embedding in film, where some world consistent with the information of the agent is required to look like the embedded film shot. This is strengthened by a normality condition. Section 5 is devoted to summation, discussion, and open issues.

2.   Possible worlds semantics for film

What is the semantics of the still photo of a spacewalk in (1)? What is the semantics of the film shot (2) of a man and a dog walking up to a door? What is the semantics of a complete film, consisting of about a thousand shots? In (1), (2), and (3), we use Scott-Strachey brackets to designate the semantic values of these pictorial artifacts. We take the approach to the semantics of pictorial materials using possible worlds semantics that was pioneered in Greenberg (2011), and developed in subsequent work. The semantics for pictures assumes a projection function π that maps a world/time w and a geometric viewpoint v in the world to a picture, $\pi(w, v) = p$. Projection is a geometric recipe, such as the perspectival procedures developed by Florentine artists in the early 15th century. Possible worlds semantics for pictures works by starting with a picture p, and interpreting it propositionally as the set of worlds that project to p from some viewpoint. Or in the alternative we employ, the semantic value of a picture p is the relation that holds between a world w and a viewpoint v iff w projects to p from v. This is stated in (4), putting the model structure that defines a class of possible worlds, and a particular projection function, outside the semantic value brackets.
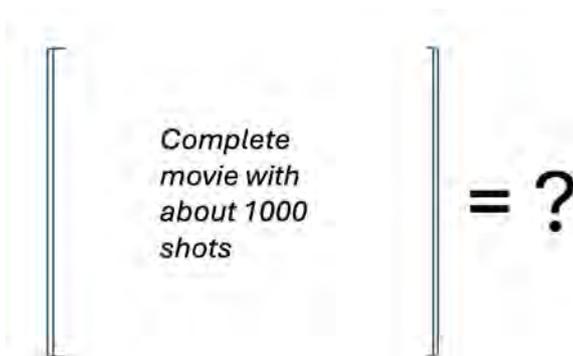
.

Over-Informative Attitudinal Embedding in Film

(1)   Semantic value of a picture


$$= \; ?$$

(2)   Semantic value of a film shot

$$= \; ?$$

(3)   Semantic value of a film

$$= \; ?$$

(4)   Relational semantic value of a picture
$$[\![p]\!]^{M,\pi} = \{\langle w, v \rangle \mid \pi(w, v) = p\}$$

Narratives such as comics that consist of a sequence of pictures have a semantics that builds on the semantics of individual pictures. (5) is a comic of two cubes moving together. Syntactically, we assume an abstract syntax that pieces pictures together into left-branching structures. By default, panel sequencing in comics has an interpretation of temporal progression, or perhaps a weaker interpretation of temporal non-regression, which allows also for temporal identity (Abusch, 2014). (7) states the non-regression option. The semantics uses variables w and w' that designate worlds at a time, and a relation of temporal extension. $w \leq w'$ is interpreted as w'

being a temporal extension of w, or possibly identical to w.[2]

(5)

(6)
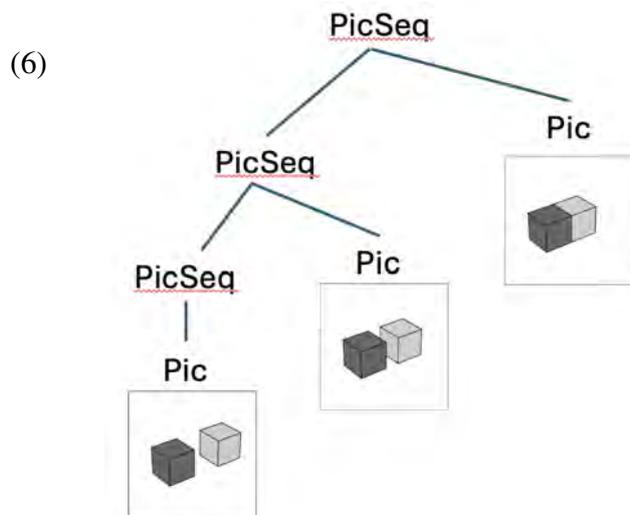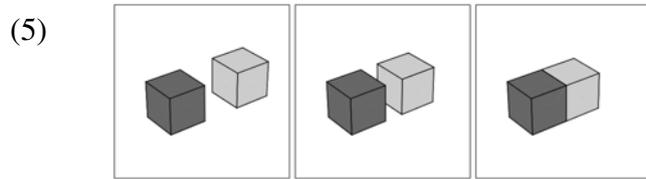
(7)

$$\left\llbracket \begin{array}{c} \text{PicSeq} \\ \diagup \diagdown \\ \text{PicSeq} \quad \text{Pic} \\ \varphi \qquad \alpha \end{array} \right\rrbracket = \left\{ \langle w', v' \rangle \middle| \exists w \exists v \begin{bmatrix} w \leq w' \wedge \\ \langle w, v \rangle \in \llbracket \varphi \rrbracket \wedge \\ \langle w', v' \rangle \in \llbracket \alpha \rrbracket \end{bmatrix} \right\}$$

(6) and (7) illustrate a general strategy for interpreting complex pictorial materials compositionally. An abstract syntax is hypothesized, and used to trigger rules of compositional interpretation. Here "abstract syntax" means that we intend to assume only that a complex piece of narrative is composed from parts, and interpreted compositionally from the interpretation of those parts. While there might be different options for the syntax, the assumption is minimal, almost amounting to the assumption that visual narratives are interpreted compositionally.

An additional aspect of the interpretation of image sequences is indexing. We would like to express that the dark cubes in the three panels of (5) are the same object, and similarly for the light cubes. For this a method for introducing indices or discourse referents for depicted objects is needed. This can be done with bounding boxes, segmentation maps, or simply geometric points within the projection of the depicted object, see (8). (9) shows a syntax for introducing an index for the light cube using a point. The idea is that the point (0.6,0.6) is within the projection of light cube, and so can be used to introduce a discourse referent for the light cube. In (10), the

---

[2]Since this is an interpretation of narrative progression, it would be reasonable to use the label Narration in (7), in place of *PicSeq*, similar to what is done in the literature on pictorial narrative progression using SDRT (Schlöder and Altshuler, 2023). By the way, the setup involving a temporal extension relation allows for branching time models.

syntax (9) is interpreted to constrain a witness for a discourse referent, using a line traced in 3D space from the viewpoint through point a in the picture plane.

(8)

| bounding box | segmentation map | point |

(9)
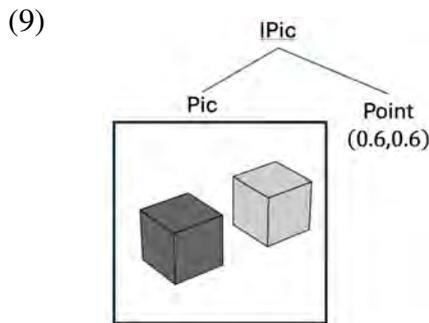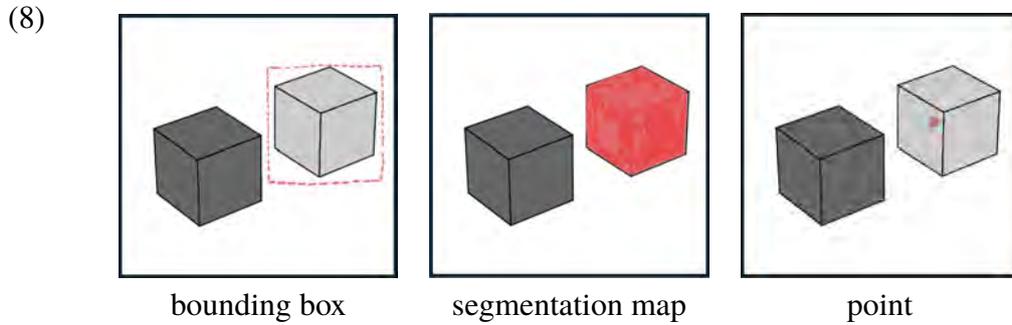
(10)   a.

$$\left\llbracket \begin{array}{c} \text{IPic} \\ \text{Pic} \quad \text{Point} \\ p \quad\quad a \end{array} \right\rrbracket = \{\langle w, v, x\rangle | \langle w, v\rangle \epsilon \llbracket p \rrbracket \wedge C(w, v, a, x)\}$$

    b. $C(w, v, a, x) \doteq$ In world/time w, object x is the first object along the directed line from v through a.

With this, a verifying tuple for a comic can be given the form $\langle w, v, \vec{x}\rangle$, where w is a world, v is the viewpoint used for the last image, and $\vec{x}$ is a vector of witnesses for discourse referents.[3] A witness for a particular discourse referent k can be looked up as $\vec{x}[k]$, and this witness is an individual in the model.

Finally, (11) illustrates a notation for coindexing. This is expressed with formal equalities between indices. This is the subtree with label Eq. This equality indicates that the light cube in the first image is the same object as the light cube in the second image. The semantics simply checks equality in the model. In the example, verifying tuples $\langle w, v, \vec{x}\rangle$, where $\vec{x}[1] \neq \vec{x}[2]$ are eliminated.

---

[3]See Abusch (2012) and Abusch (2020) on this setup for indexing, and Maier (2025) for a notation using SDRT that achieves much the same thing. The reason that v is included is that the viewpoint for the last picture is used in the semantics for introductions of discourse referents.

(11)
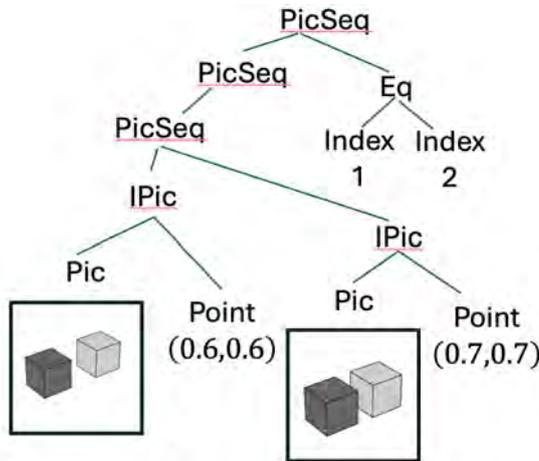
Now we extend this interpretation for image sequences to film. This is straightforward, given what was stated above. Films are composed of shots, about one thousand of them in contemporary film. An individual shot is a sequence of pictures, at a frame rate of twenty-four images per second in contemporary film. For the semantics of a shot, we use nearly the same setup as for image sequences above. A basic verifying tuple for a shot is of the form $\langle \vec{w}, \vec{v}, \vec{x} \rangle$. $\vec{w}$ is a vector of worlds, each a temporal extension of the previous one, which look like the corresponding frame in the shot from the corresponding viewpoint in the sequence of viewpoints $\vec{v}$. This sequence of viewpoints can be construed as a continuous sequence of positions for an abstract camera.

(12)   Verifying tuple for an indexed film shot or shot sequence
   $\langle \vec{w}, \vec{v}, \vec{x} \rangle$ Verifying tuple
   $\vec{w}$ Vector of worlds, each a temporal extension of the previous one
   $\vec{v}$ Vector of viewpoints, of the same length as $\vec{w}$
   $\vec{x}$ Vector of witnesses for discourse referents

Where α is a shot, a discourse referent is introduced with the syntax shown in (13). Here $\vec{a}$ is sequence of points of the same length as the number of images in α. Coindexing is implicit, in the following sense: for x to be a witness for the syntax (13), the projection of x in the ith frame of the shot must surround the point $\vec{a}[i]$, for each i. In this way, $\vec{a}$ introduces a single discourse referent, rather than multiple ones. For a human viewing a film shot, the issue of identity of objects depicted in successive frames is largely resolved by the low-level visual system (Pylyshyn, 2003). So it is reasonable to constrain the sequence $\vec{a}$ to be one which is consistent with equalities that could be inferred by the low-level human visual system. Also, since a given character may not be depicted in every frame of a shot, $\vec{a}$ must be a partial sequence, with a dummy symbol inserted in slots where the character is not depicted.

(13)   [IndexedShot $\alpha\ \vec{a}$]

The embedded passages under discussion in this paper are often composed of multiple shots, which are assembled in relations of close narration, where little or no time elapses in the described world between shots, the spatial locations described by successive shots are in close proximity, and so forth. On top of this, Cumming et al. (2017) proposed constraints on viewpoints in successive shots, such as the X-constraint. This is a formulation of the 180-degree rule on camera position, where the camera position is required to stay on one side of an "action line", such as

the line connecting two characters sitting on opposite sides of a cafe table. Since this paper does not say anything new about this, we simply assume the syntax (14a) and (14b) that combines indexed shots in a relation of close narration in a left-branching structure. Co-indexing across shots is added as in (14c), following the method of (11).

(14)   a.   [$_{\text{CloseNarration}}$ IndexedShot IndexedShot]
          b.   [$_{\text{CloseNarration}}$ CloseNarration IndexedShot]
          c.   [$_{\text{CloseNarration}}$ CloseNarration Eq]

What has been discussed so far describes extensional passages in film that are interpreted with temporal succession. There are other constructions, such as temporal flashbacks, which in the methodology employed here have to be marked with different abstract syntax labels. We analyze the passages we are discussing, such as the passage with Eileen and Randy and the passage with Nash and Charles, as composed out of their component shots as *CloseNarration*, with indexing included. The semantics of an embedded shot is then a set of tuples of the form $\langle \vec{w}, \vec{v}, \vec{x} \rangle$. Attitudinal shots must involve embedding beyond *CloseNarration*, because extensional worlds described by the films need not ever look like the embedded shots. For instance, worlds described by the film *Eileen* need not ever look like the makeout shot. Instead, extensional worlds described by the film have the character Eileen daydreaming. In the same way, extensional worlds described by *A Beautiful Mind* need not ever look like the introduction shot with Nash and Charles. Instead, they have the character Nash hallucinating. For this reason, we postulate embedding constructions such as *Daydream* and *Hallucinate* in the abstract syntax. These constructions have some additional arguments, beyond the embedded passage. One of them is indication of the character who is described as daydreaming. For the daydream passage from *Eileen*, a construal as a depiction of a daydream of Eileen's is different from a construal as a daydream of Randy's. The reason is that a given world might contain events of Eileen daydreaming, but not contain events of Randy daydreaming similar things at the same time. So for daydreams, a discourse referent should be included for the individual who is daydreaming in the extensional world of the narrative. More subtly, for Eileen to daydream about being embraced by a Randy-like individual who is in back of her is different from Eileen daydreaming that she embraces an Eileen-like individual in front of her. Therefore the syntax should include a designation of the individual in the embedded passage who is a counterpart of the daydreaming agent. (15) is our abstract syntax for an embedded shot sequence construed as depicting a daydream. *Daydream* is an abstract syntax label, and *CloseNarration* is the embedded shot. k is the index in the extensional narrative of the daydreaming agent, while j is the index in the embedded passage of the counterpart of the agent.

(15)   [$_{\text{Daydream}}$ CloseNarration j k]

In the semantics, the label *Daydream* will require the presence of daydreaming events base worlds of the narrative, the agent of which is a witness for the index k for the agent of the daydream. These events will be defined to have a semantic content. The embedded passage with label *CloseNarration* has a semantic content of the kind specified in this section. These contents need to be constrained in relation to each other. This is the topic of the next two sections.

3. The entailment problem for attitudinal shots

To get started on a discussion of the semantics of embedding, consider what it is to hallucinate a young man who looks like the one in Figure 3 saying "water". If the agent does not know he is hallucinating, the experience for the agent is the same as veridically perceiving a young man who looks like that saying "water". So a prior question is, what is it to veridically perceive a young man like that saying "water"? There is a simple model of this in epistemic event semantics. In an extensional world, there is an event, call it $e^x_{charles,water}$, of agent x looking at and hearing a Charles-like man say "water", and picking up that information. Epistemic consequences are captured by event alternatives. Since the agent picks up the information they experience, alternatives for agent x have the agent-alternative looking at and hearing a Charles-like man saying "water". We express this by saying that a similar event $e^{\bar{x}}_{charles,water}$ happens in epistemic world alternatives. $\bar{x}$ is a counterpart of the agent.

Hallucinating is not the same as veridically perceiving, so when x hallucinates, a different event $h^x_{charles,water}$ happens in an extensional world of the narrative. But the alternatives are the same, because for the agent, hallucinating is epistemically the same as veridically perceiving. Thus alternatives to a world where $h^x_{charles,water}$ happens have the corresponding veridical perception event $e^{\bar{x}}_{charles,water}$ happening.

Say an agent *eileen* daydreams that a young man like Randy embraces her from behind, and is aware that she is daydreaming. The extensional event is $d^{eileen}_{randy,embrace}$. Since Eileen is aware she is daydreaming, epistemic alternatives are not veridical perception events $e^{\bar{x}}_{randy,embrace}$, but daydream events $d^{\bar{x}}_{randy,embrace}$. But we need to capture the content of daydreaming events. This is done with independent content alternatives, which are also events, and in this case are veridical perception events $e^{\bar{x}}_{randy,embrace}$. The content alternatives have the agent alternative $\bar{x}$ seeing and feeling a young man like Randy embracing her from behind. The core part of this way of modeling perception and mis-perception in event semantics comes from dynamic epistemic logic (Van Ditmarsch et al., 2015).

Now we want to review how agent-centered natural language semantics for belief works in the framework of Lewis (1979). We refer to example (16a). Beliefs of agents are modeled with a four-place relation, with $R(w, x, \bar{w}, \bar{x})$ glossed as being $\bar{x}$ in $\bar{w}$ is consistent with x's beliefs in w. Then Eileen's beliefs in a specific world w_0 are modeled by the corresponding two-place relation $\lambda \bar{w} \bar{x}.R(w_0, eileen, \bar{w}, \bar{x})$. The relation (16b) is extracted from the embedded clause in (16a). The sentence is true in w_0 iff the two-place relation from the second step entails the two-place relation (16b).[4] That is, Eileen's agent-centered information in w must entail the agent-centered information in the complement sentence.

(16)   a.   Eileen believes a young man is embracing her from behind.
           b. $\lambda \bar{w} \lambda \bar{x}$[in world $\bar{w}$ a young man is embracing agent $\bar{x}$ from behind]

We can say that Lewis's agent-centered semantics for belief, like Hintikka's basic possible worlds semantics uses universal quantification, in that every option in the set or relation that

---

[4]For relations between worlds and individuals, relation R entails relation R' iff R represented as a set of ordered pairs is a subset of R' represented as a set of ordered pairs.

represents the agent's beliefs is required to be in the set or relation extracted from the complement sentence (Lewis, 1979; Hintikka, 1962). In linguistic semantics, it is common to use the same analysis for other attitudes, simply by starting with another four-place relation, say $H(w, x, \bar{w}, \bar{x})$ for hallucination. We are going to argue that an analysis using universal quantification is not workable for attitudinal shots in film. The reason is that embedded shots have geometric, temporal, and acoustic information that is so detailed (so informationally strong) that it is implausible that the information of an agent could be strong enough to entail it in any plausible model. We give examples of this. (17) is part of a frame in the séance scene, showing Mr. Andersen's right eye and eyebrow. There is a tuft of hair with an odd shape coming up from Mr. Andersen's eyebrow. It is implausible that Vera is being portrayed as imagining something this highly detailed. In other terms, it seems to us that in any reasonable model, there are worlds w consistent with the film as a whole, and worlds $\bar{w}$ consistent with the hallucinations of the character Vera in w, where Mr. Andersen's eyebrow in $\bar{w}$ does not look exactly like (17).

(17)

A tuft in Mr. Andersen's eyebrow as depicted in an embedded shot.

(18) is an image of the waveform of Charles's utterance of "water" in the introduction passage. This sound is part of the film shot, and is highly detailed. It is implausible that Nash is being portrayed as hallucinating anything this detailed. In fact, humans are not even capable of conceptualizing sounds at this level of detail, and Nash could not be hallucinating anything this specific in any realistic scenario.

(18)

(19) is an image of a glass pane behind Nash, taken from the introduction sequence. It has particular patterns of smudges. It is absurd to analyze the character Nash as being portrayed as hallucinating something with this specificity. But if the information of the character Nash were required to entail the content of the embedded shot, this unwelcome consequence follows.

(19)

A glass pane in Nash's lodgings as depicted in an embedded shot, with a particular pattern of smudges.

Finally, (20) shows the top of Nash's head in the hallucinated introduction scene. There is a hair projecting straight up. It is not plausible that Nash is being portrayed as hallucinating (and thus believing) that there is a hair projecting vertically from his head.

(20)

Nash depicted in an embedded shot with a hair projecting vertically from his head.

Some linguistic examples bring out the problem. (21) is a sentence that describes the introduction passage in *A Beautiful Mind*. If the information in Nash's hallucination was required to entail the content of the embedded shot, this sentence should be true, because all of Nash's hallucination alternatives according to the film would have a hair sticking up. It seems inaccurate to describe the film as attributing this belief to the character Nash.[5]

(21)   According to the movie, when Nash hallucinated meeting Charles in his grad school lodgings, he believed there was a hair projecting vertically from his head.

Example (22) exploits the fact that *would* is a universal modal. This makes *would* restricted by an *if*-clause as in (22) a near-paraphrase of a universal embedding semantics for attitudinal shots. In the shot sequence, exactly five squares of the safety glass are covered, see (23), and this is captured in the semantics of the embedded shot. We think it is incorrect to claim that the film portrays the character Eileen as daydreaming that her cheek covers exactly five squares of safety glass.

(22)   According to the movie, if what Eileen daydreamed in the second scene had transpired, things would have looked *exactly* like the shot sequence with Randy.

(23)

Eileen's cheek depicted in an embedded shot as covering exactly five squares of safety glass.

We use sentence (24) as a clue to a better analysis, with an existential modal *could* substituted. This seems like a pretty good paraphrase, or at least it is not wrong in the way that (22) is. While it is implausible to analyze the film as portraying Eileen as daydreaming that exactly five squares are covered, it is plausible to say that exactly five squares being covered is compatible with

---

[5]Sentences like (21) that describe in natural language the content of pictorial media were discussed in Ross (1997).

what she is portrayed as daydreaming. We will say that when some hidden parameters of the existential modal are set in a certain way, (24) is accurate.

(24)   According to the movie, if what Eileen daydreamt in the second scene had transpired, things could have looked exactly like the shot sequence with Randy.

Some other issues arise. Why are Nash's grad school lodgings shown as in Figure 3, for instance with a chest with five drawers? If we could say that depicting the lodgings in this way in the embedded shot has the effect of characterizing them as looking like that in extensional worlds of the film, that would be attractive.

The image in (25) is a version of a still from the greeting scene where, behind Nash, there is a giraffe peeping in from the quad. Presumably there is a world consistent with what Nash is depicted as hallucinating that looks like this from a viewpoint near Nash—he is hallucinating nothing specific about what the quad looks like outside his field of view. So why not depict the greeting sequence with a shot that looks like this? (26) is part of a still from the introduction scene, which shows Charles's necktie swinging across his chest with a particular angle and trajectory. This is also detailed and arbitrary relative to what Nash is characterized as hallucinating. What is the difference between the necktie and the giraffe?

(25)

(26)

An interim summary: (i) A semantics for embedded attitudinal shots in film using Lewis-Hintikka semantics with universal quantification attributes too much information to the agent. Maybe the quantificational force should be weakened. (ii) It would be good if embedded attitudinal shots in film contributed information about the appearance of extensional worlds of the film (the worlds where the agent is daydreaming or hallucinating) as well as attitudinal worlds. (iii) Paraphrases of the semantics of an embedded shot using a *could*-conditional are pretty good.

4.   Embedding semantics

As stated at the end of Section 2, we assume the abstract syntax for embedding shown in (27). There is an abstract syntax label *Hallucinate*, which will result in the presence of hallucination events in a verifying world. Other labels are used for other attitudes, such as daydreaming. There is an embedded indexed shot or shot sequence. On top of this, the syntax specifies the index k of the hallucinating agent in the base world. This is needed, for instance, because the interpretation of the séance sequence should distinguish an extensional situation where Vera is hallucinating from one where the medium Mrs. Ploug is hallucinating. The index j is an index of the counterpart of the agent in hallucinated situation. This is needed, for instance, because an interpretation where the character Nash takes himself to be the Nash-like man in the embedded shots is different from one where Nash takes himself to be the Charles-like man in the embedded shots. Or as stated in Section 2, an interpretation where Eileen daydreams about being embraced from behind by a Randy-like young man is different from one where Eileen daydreams about embracing an Eileen-like young woman in front of her.

(27)
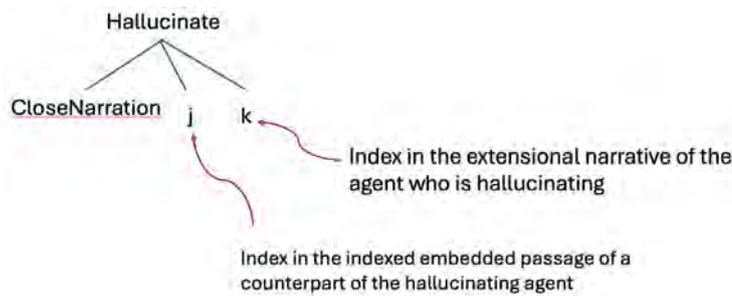
The interpretation of (27) will require that in an extensional world of the film, there are hallucination events $h_1...h_n$, the agent of which is the individual designated by the index k. Information about what is hallucinated is indirect, coming from a constraint relating the content of the hallucination events to the content of the embedded passage. Each event has a set of epistemic event alternatives (or in the case of daydreaming, content alternatives). They are highly specific, and come with preconditions for their occurrence. (28) sketches an example of what the event alternatives for hallucination events h_5 and h_9 in the Charles scenario might be. Because the alternatives are veridical perception events, they have preconditions requiring that the world and time where they happen agree with what is perceived. For instance the alternatives to h_9 can only happen in worlds where a Charles-like man is extending his right arm.

Alternative worlds are obtained by stringing together choices for alternative events, filtered by compatibility. A world of the form (29) is one where agent $\bar{x}$ first veridically perceives a Charles-like young man saying "water" with waveform f_1, and later veridically perceives a Charles-like young man extending his right arm at an angle of -5 degrees.[6]

---

[6]In the setup of Campbell and Rooth (2021), worlds are constructed as event sequences, or more precisely strings of events with interleaved guards characterizing world states. This is compatible with what we are doing here. But what we are doing here is also compatible with a setup where worlds contain token events, without being constructed from events.

(28) h_5 Agent x hallucinates a Charles-like young man saying "water".

$c_1^{\bar{x},5}$ Agent $\bar{x}$ veridically perceives a Charles-like young man saying "water" with waveform f_1.

$c_2^{\bar{x},5}$ Agent $\bar{x}$ veridically perceives a Charles-like young man saying "water" with waveform f_2

h_9 Agent x hallucinates a Charles-like young man extending his right arm.

$c_1^{\bar{x},9}$ Agent $\bar{x}$ veridically perceives a Charles-like young man extending his right arm at an angle of -5 degrees.

$c_2^{\bar{x},9}$ Agent $\bar{x}$ veridically perceives a Charles-like young man extending his right arm at an angle of -10 degrees.

(29) $... c_-^{\bar{x},1} ... c_-^{\bar{x},2} ... c_-^{\bar{x},3} ... c_-^{\bar{x},4} ... c_2^{\bar{x},5} ... c_-^{\bar{x},6} ... c_-^{\bar{x},7} ... c_-^{\bar{x},8} ... c_1^{\bar{x},9} ... c_-^{\bar{x},10} ...$

Since there are multiple event alternatives for each hallucination event, there will typically be many worlds which string together choices for event alternatives. Let $H_{\vec{h}}^{\bar{x}}$ be the set of such worlds corresponding to a vector of hallucination events $\vec{h}$. It is the centered information content of the hallucination events $\vec{h}$. In this, $\bar{x}$ is the agent counterpart, with each of the event alternatives (which are veridical perception events) required to have $\bar{x}$ as an agent. We call this the hallucination base. Roughly, we want to impose the constraint that the embedded film shot is obtained by projecting one of the worlds in the hallucination base. But this includes crazy worlds, like one where there is a giraffe peeping in through Nash's window, and one where the pane is broken. They shouldn't qualify for projection because they are not "normal" worlds. This is characterized with an ordering source, as in Kratzer's semantics for modality (Kratzer, 1978, 1981).

The following illustrates how ordering sources work. Suppose we want it to be normal for there to be a dog in the quad, normal for there to be no dog in the quad, normal for there to be no giraffe in the quad, and not normal for there to be a giraffe in the quad. It is encoded with this set of three propositions: 'there is a dog in the quad', 'there are no dogs in the quad', 'there are no giraffes in the quad'. Normal worlds are characterized by maximal consistent subsets of the set. This has the consequence that worlds with dogs are normal, worlds with giraffes are not normal, worlds without dogs are normal, and worlds without giraffes are normal. In our application, worlds with giraffes in the quad are excluded for projection to the film shot. If the proposition "the pane in Nash's window looks like (19)" is in the ordering source, all normal worlds will have the pane looking like that, as long as no other ordering propositions constrain the pane. Thus the introduction shot can be constrained to look as much as possible like Nash's lodgings, by including ordering propositions that describe the appearance of his lodgings in the extensional world of the narrative. This is a circumstantial ordering source in Kratzer's terminology. In general, the ordering source should prefer worlds that are similar to the extensional world of the film, and where nothing outrageous happens. Then, a world will be selected for projection that is consistent with the character's hallucinations, and looks as much as possible like the extensional world of the film.

Returning to the embedding construction (27), the index j for the embedded agent is used to produce an agent-centered content from the embedded indexed shot. This is the relation of a depicted agent in a world $\bar{w}$ looking like the shot from some viewpoint track, as defined in Section

2. From the embedding label *Hallucinate* and the index k in the extensional narrative of the agent y who is hallucinating, a requirement follows that agent y participates in hallucination events $\vec{h}$ in w. y is required to be the agent of these events. In the way discussed above, $\vec{h}$ determines an agent-centered content $H^{\bar{x}}_{\vec{h}}$. These are worlds where an agent $\bar{x}$ participates in events that are event alternatives to the hallucination events. In addition, we assume an ordering source $O(w)$, which includes circumstantial information about the environment of y in w.[7] Finally, we state the constraint relating the hallucination events to the embedded shot: there is some agent-centered world which a maximally normal element of the content of the hallucination events according to $O(w)$, which also satisfies the agent-centered content of the embedded shot.

In this, it is important to use the ordering source to select a world for projection *from* the hallucination content $H^{\bar{x}}_{\vec{h}}$. Then something non-normal can be depicted (such as a roommate Charles who does not exist in extensional worlds of the narrative), if all of the worlds in the hallucination content include that non-normal thing. This logic is the source of the inference that the film is characterizing Charles as being hallucinated by Nash. On the other hand, some aspects of the embedded shot, such as the number of drawers in the depicted chest of drawers, and the specific smudge pattern of window pane, can be justified by extensional worlds of the narrative looking the same way, via the circumstantial ordering source.

5. Discussion

This paper arose for us as a generalization and rethinking of the account in Abusch and Rooth (2017, 2023) of free perception sequences in comics and film, where one panel or shot shows an agent looking, and the next panel or shot shows what they see, veridically or not. In free perception sequences, the viewpoint for the second panel or shot is identified with the visual viewpoint for the experiencer—the embedded shot is a point of view shot in the strict sense. It subsequently became clear to us that, especially in film, embedded shots which do not assume the viewpoint of the experiencer are common—indeed much more common than strict point of view shots. Second, literature on embedding in film with a deep empirical base makes it clear that embedded shots in film have a range of interpretations beyond visual perception and visual hallucination, including saying and incorrect memory (Turim, 1989; Luchoomun, 2012). These points lead in the direction of positing a variety of embedding constructions marked in abstract syntax, and including in the syntax an index for the counterpart of the experiencer, as in this paper.

The earlier papers pointed out the problem of over-informativity for free perception shots, and addressed it in the same way as here, using an existential semantics and an ordering source. The difference lies in the nature of the ordering source, and the source of the base. In the earlier paper, these had to do more with properties of human perception, than general notions of normality, including circumstantial normality. We are not ready to attempt a comparison or synthesis of the two approaches, except for noting that the phenomenon is more general than free perception sequences.

---

[7]Here we follow Kratzer's formulation where ordering sources are a function from words to sets of propositions. This allows the ordering source to correspond to different circumstantial facts in different worlds consistent with the film as a whole.

The mechanism from Section 4 results in some entailments of embedded shots being in a sense lifted. In a way, something analogous comes up also for extensional shots.[8] Consider two successive shots which have a continuity mismatch that is not noticeable for any ordinary viewer, for instance a green pencil on the desk in one shot transitioning to a blue pencil in the next shot. Since no normal viewer would notice the mismatch, filmmakers do not worry about it. But a geometric semantics as employed here has described situations with a green pencil turning blue, or a blue pencil quickly substituting for a green pencil. Could this provide an alternative account of weaker information being attributed to the agent? That is, since human viewers do not extract all the information in a film shot, perhaps what is attributed to the agent is not the geometric information in the shot, but what typical human viewers would extract from the shot. Clearly, working this out involves developing a cognitive semantics for film, something that is different from the geometric possible worlds semantics employed here.

A film editor of our acquaintance, after hearing about the problem of over-informativity discussed here, commented "It's funny, we don't worry about issues like this when making a film." Filmmakers create intensional shots in the same way that they create extensional ones (whether with a camera or with a computer), and have no choice but to create intensional shots that are over-informative compared to the information of agents. In this way, the problem of over-informativity is inevitable for attitudinal constructions in the filmic medium. This is true, but it does not mean we have to give up on the project of defining a formal possible worlds semantics for film. It just means that the semantics of embedding in film is unavoidably different from the semantics of embedding in language.

The phenomenon of embedded shots in part describing the extensional situation, and in part describing the attitude of the agent, is reminiscent of de re interpretation in analyses stemming from Kaplan (Kaplan, 1969; Cresswell and von Stechow, 1982). Cresswell and von Stechow's account of informational splitting is syntactic, positing a compositional structure where part of the phrase structure coming from the embedded clause is in an extensional position, and part remains embedded. So, by analogy, we can say that the chest of drawers and the window pane in the introduction passage are interpreted de re. The analysis in Section 4 does not involve syntactic splitting, just an ordering source of a certain nature. Perhaps artfully designed ordering sources could apply also in the linguistic case, to provide a different analysis of de re readings.

References

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

Abusch, D. (2014). Temporal succession and aspectual type in visual narrative. *The Art and Craft of Semantics: A Festschrift for Irene Heim 1*, 9–29.

Abusch, D. (2020). Possible-worlds semantics for pictures. *The Wiley Blackwell Companion to Semantics*, 1–31. Circulated in 2015.

Abusch, D. and M. Rooth (2017). The formal semantics of free perception in pictorial narratives. In *Proceedings of 21st Amsterdam Colloquium*.

Abusch, D. and M. Rooth (2023). Pictorial free perception. *Linguistics and Philosophy 46*(4), 747–798.

---

[8]This was brought up in a question during our presentation at MIT.

Campbell, E. H. and M. Rooth (2021). Epistemic semantics in guarded string models. In *Proceedings of the Society for Computation in Linguistics*, Volume 4, pp. 81–90.

Cresswell, M. J. and A. von Stechow (1982). "De re" belief generalized. *Linguistics and Philosophy*, 503–535.

Cumming, S., G. Greenberg, and R. Kelly (2017). Conventions of viewpoint coherence in film. *Philosopher's Imprint 17*(1).

Greenberg, G. J. (2011). *The Semiotic Spectrum*. Ph. D. thesis, Rutgers University.

Hintikka, J. (1962). *Knowledge and Belief: an introduction to the logic of the two notions*. Cornell University Press.

Kaplan, D. (1969). Quantifying in. In *Words and objections: Essays on the work of WV Quine*, pp. 206–242. Springer.

Kratzer, A. (1978). *Semantik der Rede*. Scriptor.

Kratzer, A. (1981). The notional category of modality. *Words, Worlds, and Contexts*, 38–74.

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review 88*(4), 513–543.

Luchoomun, L. (2012). *Mental Images in Cinema: Flashback, imagined voices, fantasy, dream, hallucination and madness in film*. Ph. D. thesis, Roehampton University.

Maier, E. (2025). Pictorial language and linguistics. In *Oxford Handbook of Philosophy of Linguistics*.

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's not what you think*. MIT press.

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

Schlöder, J. J. and D. Altshuler (2023). Super pragmatics of (linguistic-) pictorial discourse. *Linguistics and Philosophy 46*(4), 693–746.

Turim, M. (1989). *Flashbacks in Film: Memory & history*. Routledge.

Van Ditmarsch, H., J. Y. Halpern, W. van der Hoek, and B. P. Kooi (2015). *Handbook of Epistemic Logic*. College Publications.