To appear in Linguistics & Philosophy.
Final pre-publication version of May 2022.

Pictorial Free Perception

Dorit Abusch and Mats Rooth

Abstract. Pictorial free perception reports are sequences in comics or film of one unit that depicts an agent who is looking, and a following unit that depicts what they see. This paper proposes an analysis in possible worlds semantics and event semantics of such sequences. Free perception sequences are implicitly anaphoric, since the interpretation of the second unit refers to the agent depicted in the first. They are argued to be possibly non-extensional, because they can depict hallucination or mis-perception. The semantics proposed here employs an account of anaphora using discourse referents, a formalized possible worlds semantics for pictorial narratives, and, to model the epistemic consequences of perceptual events, the event alternative construction from dynamic epistemic logic. In intensional examples, the second unit depicting what is seen is analyzed as embedded. It is argued that a semantics for embedding where the attitudinal state of the depicted agent is required to entail the semantic content of the picture attributes too much information to the agent. This is addressed with a model of normal looking, and a semantics for the embedding construction that uses existential quantification over alternatives, rather than universal quantification.

Keywords: Anaphora; comics; discourse representation; dynamic semantics; dynamic epistemic logic; event; film; free indirect discourse; hallucination; intensionality; perception report; pictorial narrative; possible world; premise semantics.

1. Introduction

Free perception sequences in our sense are ones where a setup sentence in language, a panel in comics, or shot in film shows or describes a character looking, and a subsequent sentence, panel, or shot is understood as describing or showing what the character saw when looking. In (1a), from S. Tan's *The Arrival*, one panel shows a man looking down, and the other shows some strange writing on the sidewalk (Tan, 2006). It is understood that the second panel shows what the man

This paper is part of the Special Issue "Super Linguistics" of Linguistics & Philosophy, edited by Pritty Patel-Grosz, Emar Maier and Philippe Schlenker. Versions of parts of the material were presented at the 21st Amsterdam Colloquium, Sinn und Bedeutung 25, and Sinn und Bedeutung 26, and were published in Abusch and Rooth (2017), Abusch and Rooth (2021), and Abusch and Rooth (2022). Thanks to the audiences for their reactions. Thanks to Helena Aparicio, Forrest Davis, Jon Ander Mendia, Marten van Schijndel, Yexin Qu, and Ede Zimmermann for numerous comments and examples. The paper in its late form was discussed in Semantics Seminar at Cornell in Fall 2021. Warm thanks to the participants for their reactions. Three L&P reviewers and Emar Maier as editor provided valuable comments, which are reflected in the final paper. Errors are our own. Thanks to Milka Green for assistance with artwork. Images that are quoted from comics and film are used for educational and critical purposes, and are property of their respective owners.

sees. (1b) is from Simone Lia's *Fluffy* (Lia, 2007). The character Michael has lost his rabbit on a train, and searching, he looks into a cabin and hallucinating, sees a girl with a rabbit in a sandwich.

(1) a.   b.  

The same phenomenon is found in film. (2) shows three frames from *The Third Man* (Reed, 1949). The character with the fur collar seen in the first two frames looks up, and the subsequent shot shows a man approaching a woman while taking off his hat. It is understood that this is what the man in the first shot sees. In film terminology, the second showing what that character sees is a point of view shot.[1]

(2) From *The Third Man* (Reed, 1949). Baron Kurtz (Ernst Deutsch), Dr. Winkel (Erich Ponto), Holly Martins (Joseph Cotten), Anna Schmidt (Alida Valli).

(3a-d) are linguistic examples.[2] In (3a-c), the first clause describes a character looking, and the second clause describes what the character sees. The clauses are linked by a conjunction in the first case, a when-clause in the second, and are sequenced at the discourse level in the third. (3d) is less direct, in that the first clause describes the character turning around, rather than looking. It

[1]Branigan (1984) describes a sequence of two shots A and B as follows.

Shot A:	Point/Glance	
	1. *Point:*	establishment of a point in space.
	2. *Glance:*	establishment of an object, usually off-camera, by glance from the point.
Between Shots A and B:		
	3. *Transition:*	temporal continuity or simultaneity.
Shot A:	Point/Object	
	4. *From Point:*	the camera locates at the point, or very close to the point, in space defined by element one above.
	5. *From Point:*	the object of element two above is revealed
Shots A and B:		
	6. *Character:*	the space and time of elements one through five are justified by – referred to – the presence and normal awareness of a subject.

Depending on the interpretation of "glance" and item 6, this potentially allows for sequences where the viewing character is not depicted in Shot A.

[2](3a) is from a report by Larry Gross in CityBeat (Gross, 2006). (3b) is from the story "Ghosts" by Brian Hart (Hart, 2014). (3c) is from Lawrence's *Sons and Lovers*, as quoted by Brinton (Lawrence, 1913). (3d) is from Brinton (1980).

seems to be accommodated that the character looked, with the final clause describing what he sees. Examples like these are discussed in Brinton (1980), who relates them to the larger phenomenon of free indirect discourse (Banfield, 1982; Sharvit, 2008; Eckardt, 2014). Hinterwimmer (2017) argues that (3b), while in some respects similar semantically to free indirect discourse, is a distinct construction he calls viewpoint shifting.[3]

(3) a. I looked back up the sidewalk, and that angry kid was walking toward me.
b. When I looked up a guy with a metal detector was walking toward me.
c. He looked at his mother. Her blue eyes were watching the cathedral quietly.
d. "Look!" Fred turned around. Jack was coming across the street towards him.

This paper analyzes pictorial free perception sequences semantically. The framework combines possible worlds semantics, dynamic semantics and event semantics. The semantic value of a picture is defined to be a certain kind of centered proposition, as modeled in possible worlds semantics. To theorize about anaphoric co-reference across pictures in a narrative, discourse referents are introduced. Reasoning about the epistemic consequences of looking uses event semantics.

One concern in the paper is to develop in a dynamic framework the idea that free perception constructions are implicitly anaphoric, in that the interpretation of the second panel or shot refers to the agent that is depicted in the first panel or shot. A second concern is to develop in a possible worlds and event framework an analysis of intensional sequences such as (1b). This turns out to be non-trivial for reasons that are particular to pictorial materials. In linguistic and philosophical literature, it is argued that sentences with bare infinitive complements as in (4a) are not intensional in the same way as (4b) with a tensed complement (Barwise, 1981; Higginbotham, 1983; Felser, 1999). In the final version of our final analysis, pictorial free perceptions sequences (including intensional ones) are in important respects parallel to linguistic bare complements rather than linguistic tensed complements.[4]

(4) a. Fred saw Jack come across the street toward him.
b. Fred saw that Jack was coming across the street toward him.

The paper is organized like this. Section 2 reviews the semantic framework from Abusch (2012a, 2020); Rooth and Abusch (2019), which is a pictorial discourse representation theory, and provisionally introduces discourse representations for extensional and intensional free perception. Section 3 develops a detailed semantics and model theory for pictorial free perception, in a framework where the epistemic consequences of perceptual events are modeled using event-alternative relations. The initial analysis of intensional free perception is parallel to a Hintikka analysis of intensional clausal embedding in natural language, where the epistemic state of the agent is required to entail the content of the embedded material. Section 4 brings up a problem with this: it is implausible that agents in described situations should have strong enough attitudinal information to entail the detailed geometric information (and for film, temporal information, and for film with

[3]As pointed out by an L&P reviewer, (3a,b) are parts of first-person narratives, and thus potentially have special properties related to the narrator and the pragmatic status of the narrative as a whole. While this is true, the examples have the same interpretation implying perception when they are transposed to third person.

(i) He looked back up the sidewalk, and that angry kid was walking toward him.
(ii) When he looked up, a guy with a metal detector was walking toward him.

[4]The linguistic parallels are important in Section 4, where we show that our initial analysis of intensional pictorial free perception from Section 3 wrongly delivers results parallel to the tensed (4b).

sound, acoustic information) of pictures and film shots. The problem is addressed in Section 4 by introducing a model of the normal epistemic consequences of perceptual events, and in Section 5 by weakening the quantificational force of the intensional free perception construction. Section 6 takes stock of the results of the investigation.

2. LFs for free perception

This section proposes logical forms or discourse representations for free perception sequences. The framework is the one presented in Abusch (2020), where a linear pictorial narrative is augmented with introductions of discourse referents, equalities between discourse references that express anaphora, and unpronounced operators.[5] These structures are comparable to "logical forms" or "compositional structures" for natural language sentences, in that they have similar shape to surface sentences, but have indexing resolved, and have certain operators inserted that are not present in the pronounced sentences. They are structures of a syntactic nature that are interpreted in possible worlds semantics in a deterministic way. To take a linguistic example, it is sometimes maintained that the English sentence (5a) has an LF (logical form) like (5b), where an unpronounced generic operator GEN (which is comparable to an adverb of quantification like *always*) has been inserted, the subject [female spiders] has been moved to the position of the restriction of GEN, and binding operator (the index 1) has been inserted that binds the empty category e_1 and the pronoun [their$_1$]. The result of these operations is the structure that is interpreted compositionally.[6] In parallel fashion, we assume LFs for pictorial narratives that are similar to surface comics or surface films, but have added indexing and added operators.

(5) a. Female spiders eat their mates.
b. [GEN [female spiders] [1 [e_1 eat [their$_1$ mates]]]]

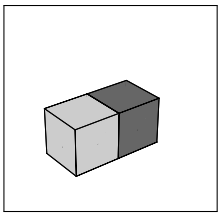
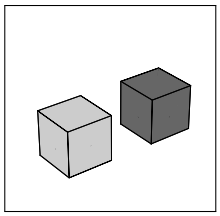
The structures introduced below can alternatively be viewed as discourse representation structures in the sense of Discourse Representation Theory (Kamp, 1981; Kamp and Reyle, 1993), with a syntax for discourse referents and for equalities between them that captures anaphora. In place of the box syntax of classical DRT, we use a linear notation that is adapted from Predicate Logic with Anaphora, PLA (Dekker, 1994, 2012), which has the virtue of being notationally concise.[7]

To illustrate the basic notation, (6) is a comic of a dark cube moving away from a light cube, including information in discourse referents that identifies the light cube in the second panel with the light cube in the first one, and the dark cube in the second panel with the dark cube in the first. A basic pictorial narrative is defined as a sequence of pictures. The LF is formed by interleaving syntax related to indexing.

[5]While this paper aims at being self-contained, the handbook article Abusch (2020) has motivational examples and more detailed explanation of the dynamic and possible worlds frameworks. Greenberg (2011) introduces and motivates the possible-worlds analysis of pictorial contents. Readers who wish to see general introductions to the semantic framework should refer to these.

[6]This is a tripartite logical form like the ones discussed with various applications in Lewis (1975), Kratzer (1978), and Heim (1982).

[7]Nothing much hinges on this choice, but we do think that Dekker's version is the most elegant formulation of dynamic semantics, and follow it for that reason, in addition to notational simplicity.

(6)  $(0.3,0.3),(0.6,0.4)$  $(0.3,0.3),(0.7,0.5),1=3,\ 2=4$

The panels in the example are assumed to be unit squares. A tuple such as $(0.3,0.3)$ is construed as a point in the unit square, which refers to the previous pictures and serves to introduce a discourse referent, in this case a discourse referent for the light cube. This comes about because the point $(0.3,0.3)$ is within the projection in the picture plane of the light cube in a described situation for the first picture.[8]

PLA uses a recency convention for indexing, where 1 is used for the most recently introduced discourse referent (henceforth dref), 2 is used for the penultimately introduced dref, and so forth.[9] This results in the interpretations given in (7) for the four drefs introduced by the notation in (6).

(7)

dref	interpretation
1	dref for the dark cube (depicted in) the second panel
2	dref for the light cube in the second panel
3	dref for the dark cube in the first panel
4	dref for the light cube in the first panel

The equalities in (6) are formal equalities between discourse referents, which have a semantics of identity in the model. The equality "1=3" indicates that the dark cube depicted in the second picture is identical to the dark cube depicted in the first picture. This will be phrased more carefully in a moment.

Greenberg (2011) introduced a program of giving pictures a possible-worlds semantics, comparable to the possible worlds semantics for natural languages given in Montague Grammar and its descendants. He proposed defining semantic values of pictures using geometric projection functions, such as perspective projection or orthographic projection. In the formulation assumed here, the semantic value of a picture q, written $[\![q]\!]^M$ is a relation between worlds and viewpoints. A viewpoint is an oriented location in space, essentially a location in space accompanied by information that locates an oriented picture plane. This is comparable to the station point in classical formulations of perspective rendering (Alberti, 1435). Where w is a world and v is a viewpoint, $\pi(w,v)=q$ indicates that world w projects to picture q from viewpoint v, using e.g. linear perspec-

[8] In the place of points, one can use areas in the picture or bounding boxes for the introduction of discourse referents. See Abusch (2012a) and Abusch (2020) for some discussion. In machine vision research, bounding boxes are used as proxies for depicted individuals, in effect introducing discourse referents (Liu et al., 2020). Abusch (2012a) introduced geometric discourse referents for depicted individuals in comics, and the scheme for anaphora using formal equalities in the LF. It uses areas in a picture rather than points to introduce drefs.

[9] While the recency convention is a convention, it is one which is intertwined with the semantics of formulas of PLA, and of formulas such as (6), which is defined in a way that the recency convention is observed (Dekker, 1994, 2012; Abusch, 2020). Bittner (2001) proposed re-configuring recency into a model of salience or centering, which is also applied in compositional semantics. For instance, in her analysis, "topic chains" in Mandarin keep constant a dref referenced with the index 1, or in her notation the term $\top_1$. Centering is potentially relevant for pictorial narratives, since a depicted individual can be more or less salient. This is not incorporated in the formal model used here. Thanks to an L&P reviewer for pointing out the relevance of salience of discourse referents in pictorial materials.

tive. π is the projection function. Viewpoint-centered semantic values are obtained by inverting projection, $[\![q]\!]^M = \{\langle w,v\rangle | \pi(w,v) = q\}$.[10] Informally, the semantic value of a picture is the set of all pairs of worlds and viewpoints such that the world looks like the picture from the viewpoint. The superscript M outside the semantic value brackets is a model structure defining the space of possible worlds.[11]

(8) *Viewpoint-centered propositional semantic value for a picture*
$[\![q]\!]^M =_{\text{def.}} \{\langle w,v\rangle | \pi(w,v) = q\}$

The worlds under discussion here are worlds at a time. Where w and w' are worlds, $w \leq w'$ indicates that w' is a temporal extension of w. This is consistent with branching-time models, where a world w can have extensions w' and w'', where w' and w'' are not comparable.[12] The semantics of narratives such as (6) is covered by a satisfaction definition in the format $w,v,\mathscr{O} \models \Psi$, where Ψ is the pictorial narrative, w is a world, v is a viewpoint and $\mathscr{O}$ is a tuple of witnesses for discourse referents. Satisfaction is defined inductively. For instance, (9) is the clause covering a narrative Ψ extended with an equality between discourse referents. The natural numbers n and m function as pronouns that take values by indexing into the sequence of witnesses $\mathscr{O}$. The second premise checks equality in the model.

(9) *Incrementing with an equality*

$$\frac{w,v,\mathscr{O} \models \Psi \qquad \mathscr{O}[m] = \mathscr{O}[n]}{w,v,\mathscr{O} \models \Psi\ m{=}n}$$

The viewpoint v in a satisfaction statement $w,v,\mathscr{O} \models \Psi$ has the interpretation of the viewpoint used to project the last picture in Ψ. This comes about from the semantic rule (10), which defines the satisfaction condition for a narrative Ψ incremented with a picture q. The premise $w,v,\mathscr{O} \models \Psi$ says that $\langle w,v,\mathscr{O}\rangle$ satisfies the pictorial narrative Ψ. The premise $w \leq w'$ optionally advances time (Abusch, 2014). The premise $\pi(w',v') = q$ says that world w' looks like picture q from viewpoint v'. The viewpoint is memorized in the output statement $w,v',\mathscr{O} \models \Psi\, q$. $\mathscr{O}$ is unchanged, because there is no new discourse referent.[13]

[10] Viewpoint-centered semantic values are analogous to the agent-centered semantic values in Lewis's de se analysis of attitudes (Lewis, 1979). But the two are distinct, since a world can be depicted from a viewpoint without there being an agent there, and some agents (such as distributed AI systems without any vision apparatus) do not correspond to viewpoints.

[11] As pointed out by an L&P reviewer, in the definition (8) , the metavariable q for pictures occurs on the right hand side of the equation as well as the left. This differs interestingly from lexical semantic definitions for natural language such as

$[\![\text{chien}_{\text{N}}]\!] = \{\langle w,x\rangle | \text{DOG}(w,x)\}$,

where DOG is a relation symbol in the metalanguage, and *chien* does not occur on the right hand side. On both sides of (8), q is a picture, which can be constructed e.g. as a function from the discrete 512×512 square to grayscale values. The difference between the definitions comes from the fact that the semantic value of pictures is defined substantively or "iconically", in a way that is sensitive to their geometry.

[12] The event-sequence models used in Sections 3-5 are branching-time models. Thomason (1984) describes branching-time models as they are used in linguistic semantics and the philosophy of language.

[13] See Abusch (2020) for more on this. The satisfaction relation is defined inductively, using clauses stated in natural deduction form, with premises and a conclusion. The defined satisfaction relation is the smallest relation that obeys the clauses.

(10) *Incrementing with a picture*

$$\frac{\begin{array}{l} w, v, \mathscr{O} \models \Psi \\ w \leq w' \\ \pi(w', v') = q \end{array}}{w, v', \mathscr{O} \models \Psi\, q}$$

Finally, (11) is the clause for introduction of a discourse referent. It traces a line from the viewpoint v through the point a, and defines the witness for the new discourse referent to be the first object along that line in w. Informally, the new witness y is an object in w such that point a is within the depiction of y in the last picture of Ψ.[14] This definition enforces the recency convention, since in the new tuple of witnesses $\mathscr{O}y$, the new witness is referenced with the index 1, $\mathscr{O}y[1] = y$.

(11) *Incrementing with a dref*

$$\frac{\begin{array}{l} w, v, \mathscr{O} \models \Psi \\ \text{The minimal point on the directed line from } v \text{ through } a \\ \text{that is in an object in } w \text{ is in the object } y \text{ in } w. \end{array}}{w, v, \mathscr{O}y \models \Psi\, a}$$

This completes the review of a dynamic possible worlds semantics for pictorial narratives enriched with discourse referents and equalities among them. The definition makes available a semantic value of a narrative as a relation between worlds, viewpoints, and tuples of witnesses for discourse referents. And it makes it possible to refer to depicted individuals by indexing into the witness-sequence.[15]

Pictorial free perception sequences are anaphoric, in that the interpretation of the second picture refers back to an agent depicted in the first picture. Roughly, in the free-perception sequence (1a), the world is claimed to look like the second picture from the visual-geometric perspective of the man touching a hat who is depicted in the first picture. In the framework developed here, anaphora is made explicit by the introduction of a discourse referent, and of an operator that refers to that discourse referent. In (12), $(0.5, 0.5)$ is a geometric point that introduces a discourse referent for the man looking down while holding his hat.[16] L_1 is a constant symbol that introduces extensional free perception semantics. The semantics of L_1 refers to the individual $\mathscr{O}[1]$, which is the witness for the most recently introduced discourse referent.[17]

[14]This combines random selection of a new witness y with a geometric test on y, the viewpoint v, the geometric point a which is a syntactic part of the narrative, and world w. There is no explicit reference to the previous picture, but other parts of the definition ensure that w looks like the last picture in Ψ from v.

[15]Rooth and Abusch (2019) apply the same framework to cross-modal indexing. Maier (2019) and Maier and Bimpikou (2019) theorize about speech balloons in comics, and about attitudes such as imagination in comics, using a DRT framework with box syntax. A version of Abusch (2020) circulated in 2015 defined the dynamic framework as applied here.

[16]This comes about because the point is within the projection of the man.

[17]Thus we use a constant L_1, rather than formulas like L_i, where i is a natural number. Since we can choose what dref is introduced last, this does not reduce expressiveness. Nevertheless, the "1" in L_1 is a reminder that the agent is the individual picked out with the index 1.

(12)  (0.5,0.5)  L_1

The general semantics for pictorial narratives already entails that if $\langle w, v, \mathscr{O} \rangle$ satisfies a version of formula (12) without L_1, then w looks like the sidewalk picture from viewpoint v. To enforce a free-perception semantics, it has to be added that v is the visual-geometric viewpoint of the agent $\mathscr{O}[1]$, i.e. of the agent depicted in the first picture. We go a bit further. We want to capture the fact that in pictorial free-perception sequences, the agent in the setup picture is shown explicitly as looking. Loosely, the content of the free-perception construction is "things look like the second picture from the visual-geometric perspective of the agent in the first picture who is depicted as looking". This suggests including in the semantics of L_1 a presupposition that the agent $\mathscr{O}[1]$ is looking. In order to satisfy the presupposition, the agent $\mathscr{O}[1]$ that is set up by the geometric discourse referent should appear to be looking. Often in film, the setup shot makes it very obvious that a certain character is looking, by showing the character looking up or glancing.[18] We can say this facilitates a free perception interpretation, by making available a discourse referent and the salient information that the witness for the discourse referent in a described situations is looking, thus satisfying the presupposition of L_1. Thus we propose the semantics in (13).

(13) *Semantics of the extensional free perception constant L_1, referring to a satisfying tuple $w, v, \mathscr{O}$*

Presupposition: $\mathscr{O}[1]$ is an agent with a visual system who is looking in world w.
Assertion: v is the visual-geometric viewpoint of $\mathscr{O}[1]$.

The free-perception construction as formalized using L_1 is technically extensional, in the sense that the semantics (13) refers only to the base world. Acts of looking have intensional consequences though, because they allow agents to pick up information. Roughly, an agent who looks at a sidewalk will have epistemic alternatives where the counterpart of the agent is looking at a similar sidewalk. For this reason, a formally extensional LF for free perception can have intensional entailments. This fact is developed in Section 3, where we describe possible worlds models which provide a model of the epistemic consequences of looking.

The analysis so far has packed free-perception semantics into a constant L_1, which combines conjunctively with the rest of the visual narrative. This syntax is in one way too unconstrained. Consider the narrative and satisfaction sequence (14) with three pictures, one discourse referent introduced after the first picture, and an L_1-operator after the third picture. The geometric discourse referent is a, which is understood to be a point within the depiction of the man in the fur collar in the first picture.[19] Because of the syntax, the semantics coming from L_1 is that the viewpoint v for the third picture is the visual viewpoint of the agent shown in the first picture. This has the third picture rather than the second one showing what the man in the first picture sees. The problem comes from no discourse referent being introduced after the second picture. We rule this out in the syntax, by introducing the operator L_1 with the phrase structure rule (15b). It introduces L_1

[18] This is true in example (2) from *The Third Man*.

[19] We also use the notation a in later examples, rather than showing the point as a specific ordered pair.

Ψ	Meta-variable for pictorial narratives.
q	Picture, a part of a pictorial narrative.
x	Geometric point that introduces a new discourse referent, a syntactic part of a a pictorial narrative. In Ψx, index 1 is the new discourse referent.
$\mathscr{O}$	Tuple of witnesses for discourse referents.
$m = n$	Equality between discourse referents.
w	Possible world at a time.
$w \leq w'$	World w' is the same as w, or is an extension of it.
v	Geometric viewpoint.
π	Projection function.
$\pi(w,v) = q$	Picture q is the projection of w from v.
Q	Pictorial content, a relation between worlds and viewpoints.
$Q(w,v)$	Viewpoint-centered world $\langle w,v\rangle$ satisfies pictorial content Q.
$V(w,x)$	Geometric visual viewpoint in world w of agent x.
$w,v,\mathscr{O} \models \Psi$	Tuple $\langle w,v,\mathscr{O}\rangle$ satisfies narrative Ψ.
L_1	LF syntax for extensional free perception.
$S_1(q)$	LF syntax for intensional free perception.

Figure 1: Notation for the syntax and semantics of pictorial narratives.

along with an immediately preceding picture and discourse referent, in place of the unconstrained syntax (15a). In these context-free rules, Ψ is the syntactic category of pictorial narratives, D is the syntactic category of geometric discourse referents, and q is the syntactic category of single pictures.[20]

(14)

$w,v,\mathscr{O} \models$  a   L_1

(15)a. $\Psi \rightarrow \Psi\ L_1$

b. $\Psi \rightarrow \Psi\ \mathrm{D}\ q\ L_1$

This completes the initial development of an extensional logical form for free perception examples. Figure 1 summarizes some of the notation. We turn now to the LF of examples like (1b), which are understood as portraying hallucination or mis-perception. The syntax of this example

[20] An L&P reviewer feels that the sequence of images in (14) can attribute seeing the scene in the third image to the man depicted in the first image. We would have to be convinced by examples that this is possible in film and comics. As also pointed out by a reviewer, "the point of view shot can come before the shot that depicts the person looking (i.e. the order of the sequence can be reversed, while still being parsed as a POV connection", with examples presented in Branigan (1975). In the framework used here, this calls for the syntax $q\ \hat{L}\ p\ a\ 1 = 2$, with $\hat{L}$ introducing a discourse referent rather than picking one up. We are not confident enough about data in film and literature in film theory to defend a specific proposal. In the methodology pursued here, using a formalized syntax and semantics, the place to state the right empirical generalizations about point of view constructions is in the syntax of pictorial narratives.

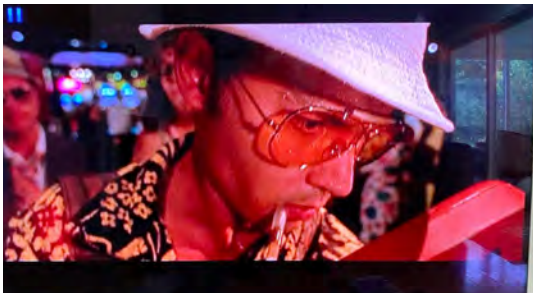
is not as in (16), because that would entail that the base world of a described situation has a girl eating a rabbit sandwich, or looks like a girl is eating a rabbit sandwich. To take another example, in a passage from the film *Fear and Loathing in Las Vegas* the protagonist looks down, and in a state of drug-induced hallucination, sees the feet of the man in front of him and adjacent furniture covered in paisley patterns (Gilliam, 1998).[21] The syntax of this example is not as in (17), because that would entail that the base world of the described situation for the film has feet and furniture covered with paisley patterns. This is not how the narratives are understood. In *Fear and Loathing in Las Vegas* sequence, described situations should have the protagonist facing a man with lower limbs of normal appearance, while it is alternatives for the protagonist that have a man with paisley patterns on his lower limbs.[22]

(16)

a

L_1

(17) From *Fear and Loathing in Las Vegas* (Gilliam, 1998). Duke (Johnny Depp).



a

L_1

These considerations indicate that, in the framework we are assuming, the second picture in a free-perception sequence that describes mis-perception or hallucination is not a top-level conjunct. For if it were, the content of that picture would be an entailment of the narrative. Taking a lead

[21] Thanks to Marten van Schijndel for this example.

[22] There is actually a competing analysis of *Fear and Loathing in Las Vegas*, where the entire movie describes worlds consistent with the perceptions and information of the protagonist, rather than "realistic" base worlds where there are no paisley patterns crawling up the legs of characters. In this competing analysis, the two frames in (17) are in the same conjunctive context, and the passage is an example of extensional free perception. The competing analysis is suggested by the novel version of *Fear and Loathing and Loathing in Las Vegas*, where a description of hallucination begins with the second sentence of the book (Thompson, 1972).

> We were somewhere around Barstow on the edge of the desert when the drugs began to take hold. I remember saying something like "I feel a bit lightheaded; maybe you should drive ..." And suddenly there was a terrible roar all around us and the sky was full of what looked like huge bats, all swooping and screeching and diving around the car, which was going about a hundred miles an hour with the top down to Las Vegas.

This suggests a reading where most (or all) of the novel and film, including the setup shot in (17), reflect the perception and information of the protagonist. Different versions of this point about *Fear and Loathing in Las Vegas* are made in Abusch (2022) and Liefke (2022), commenting on Maier (2022).

from the syntax of attitudinal embedding in language, we propose that these examples have an embedding predicate at the level which is semantically interpreted, which we write as S_1 (for see). Where q is a picture, the syntax is $S_1(q)$. As with L_1, the subscript 1 in S_1 is a reminder that the agent is the individual picked out by index 1. So instead of (16), the LF of the passage from *Fluffy* is (18).

(18)  $a\ S_1$ 

In a satisfying tuple $\langle w, v, \mathscr{O} \rangle$ for this sequence, $\mathscr{O}[1]$ is the man depicted in the first picture. In the semantics of $S_1(q)$, the individual $\mathscr{O}[1]$ will be the agent who sees a scene like q. This paper will develop several options for an intensional semantics for this syntax, beginning here with a Hintikka semantics involving universal quantification over epistemic alternatives (Hintikka, 1962). We use the notation $R(w,x,u)$ for world u being an alternative for an agent x in world w. [23]

For orientation, (19) defines a semantic attitude predicate A that combines with a world w, an individual x (the subject of the attitude) and a proposition p (contributed by a complement sentence). The definition refers to a world alternative relation R, which in the formula $R(w,x,u)$ is read as "world u is an alternative for agent x in world w". The definition says that A holds between x and p in world w if and only if every world that is an alternative for x in w satisfies the proposition p.

(19) $A(w,x,p) \equiv_{\text{def.}} \forall u[R(w,x,u) \rightarrow p(u)]$

There is a mismatch between (19) and the semantics of pictures, because on our assumptions, a picture contributes a relation between worlds and viewpoints, rather than a set of worlds. This is remedied by introducing reference to the visual-geometric viewpoint of the agent of the attitude. Let V be the function that maps a world w and an agent x who exists in w to the visual-geometric viewpoint of x in world w. (20) adapts (19) to a pictorial complement by finding the visual-geometric viewpoint $V(u,x)$ of agent x in the alternative world u. S is the attitudinal predicate being defined. Relation S holds for world w, agent x, and pictorial content Q if and only if for every alternative u for x in w, world u and the visual-geometric viewpoint of x in u satisfy the pictorial content.[24] R^S is the alternative relation that is appropriate for descriptions of seeing.

(20) $S(w,x,Q) \equiv \forall u[R^S(w,x,u) \rightarrow Q(u,V(u,x))]$

With this, (21) defines the semantics of the piece of syntax $S_1(q)$ in a visual narrative, where q is a picture.[25] It applies the relation S defined in (20) to the semantic value of picture q and to the agent $\mathscr{O}[1]$. In words, it is required that for every alternative u for the agent $\mathscr{O}[1]$ in world w, world

[23]Or in a de se semantics, $R(w,x,w',x')$ means that the centered world $\langle w',x' \rangle$ is an alternative for x in w. We use a non-centered semantics, because a general de se development of event sequence models is non-trivial.

[24]This uses identification of x across worlds, and so is problematic given Lewis's argument for agent-centered modeling of epistemic attitudes (Lewis, 1979). The de se version of the definition is $S(w,x,Q) \equiv \forall w' \forall x'[R(w,x,w',x') \rightarrow Q(w',V(w',x'))]$, which does not use identification of individuals across worlds.

[25]This definition is part of the inductive definition of the satisfaction relation.

u looks like picture q from the visual-geometric viewpoint of agent x in u. In the definition, Ψ is a pictorial narrative that gets incremented to $\Psi\, S_1(q)$ by adding a picture embedded under a seeing predicate.

$$(21)\quad \frac{\begin{array}{l} w, v, \mathscr{O} \models \Psi \\ S(w, \mathscr{O}[1], \lambda w' v' . \pi(w', v') = q) \end{array}}{w, v, \mathscr{O} \models \Psi\, S_1(q)}$$

What is the alternative relation R^S that is involved in this analysis? One could reasonably maintain that there is a special-purpose relation of being a visual alternative for an agent. In the model-theory developed in Section 3, the relation will instead be the ordinary epistemic alternative relation. Semantic effects that are particular to visual perception will be expressed with a presupposition that the agent is involved in a looking event, together with reference as in (20) to the visual-geometric viewpoint of the agent.

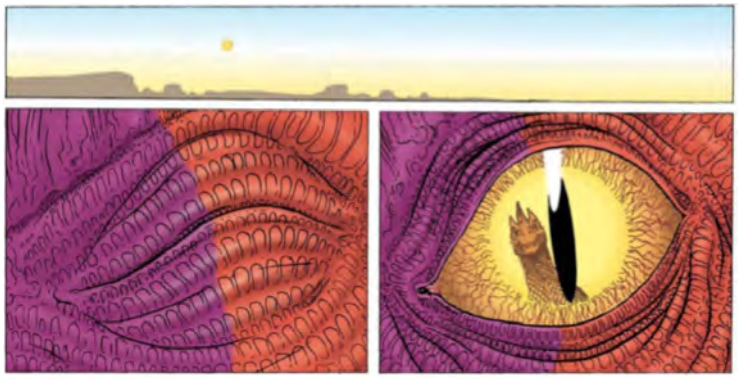
So far, this section has proposed two logical forms for pictorial free perceptions sequences. The extensional or quasi-extensional one using the constant L_1 enforces a coincidence between the viewpoint for the second picture with the visual-geometric viewpoint of the agent depicted in the setup picture. This is augmented with a presupposition that the agent is looking. The intensional LF using S_1 involves embedding of the second picture, and this has the semantics of clausal embedding, with an adaptation to the type of the pictorial complement. In the framework that is assumed, free-perception sequences that are understood as portraying mis-perception must involve embedding of the second picture. At the current point in the argument, we have not shown whether examples that are understood veridically, such as example (1) with the man looking at the sidewalk, and example (2) from *The Third Man*, have the embedding syntax (using S_1), or the non-embedding syntax (using L_1). It could be attractive to use S_1 uniformly, and to posit that what amounts to an extensional reading is obtained by accommodating that the agent is looking veridically. But there are some arguments by analogy for the extensional syntax.

(22) is from Cece Bell's *El Deafo*, and shows the heroine Cece and another character, Ginny (Bell, 2014). A dotted sightline indicates that Ginny is looking at Cece's hearing aid. Sightlines are a convention that indicate the visual focalization of a depicted agent. The information that is conveyed is similar to what would be conveyed by a free perception sequence, with one panel showing Ginny looking, and the next panel showing the hearing aid and the top of Cece's head. The information conveyed by (22) appears to be entirely extensional—the characters are in a certain geometrical configuration, and Ginny is visually focalizing on a certain point. The panel carries the information that Ginny is looking, and gives information about what she is focalizing visually. Arguably, the intensional implication of the panel—that Ginny sees what the hearing aid looks like, and by implication is puzzled about it—follows from the extensional content of the panel.

(22)

The three-panel sequence (23) is from Delgado's *the Age of Reptiles* (Delgado, 1997). A predatory dinosaur opens its eye, and in the last panel, another dinosaur is seen reflected in the eye. It is inferred that the predatory dinosaur sees the other one, with an ominous implication that it has spotted its prey. But the literal information in the panel is extensional.

(23)



Consider for a moment what would be involved in an intensional syntax and interpretation for (23). The panel would have to be broken down into two sub-panels, one showing the dinosaur looking, and another a small, syntactically embedded subpanel showing what is seen. This amounts to a "vision bubble" embedded in an image of the agent's eye. There are pictorial bubbles such as example (24) from Bilal's *Froid Équateur* (Bilal, 1992). But a bubble division is otiose in the case of (23), because of the possibility of a straightforward syntax and extensional interpretation.

(24)

(25) is from Tezuka's *Ode to Kirihito* (Tezuka, 2015). It shows a hulking figure at a door, with his head tilted down towards the hero Kirihito on the floor. It can be inferred that the hulking figure

sees a view approximately like the part of the panel surrounding Kirihito. But the panel as a whole could not show what the hulking character sees, because he himself is depicted, and he could not see himself. Here again an extensional analysis is attractive.[26]

(25)

These three panel types (with sight lines, eye reflections, and over-the-shoulder viewpoint) are informationally similar to veridical free perception. There is little temptation in these cases to formulate an intensional analysis based on a syntax with embedding, since the inferences that readers make about what characters see are supported by the extensional content of the panel. This tends to favor an extensional analysis of veridical free perception, because there too (assuming a switch in geometric viewpoint as enforced by L_1), the inferences that readers make about what the agent sees are supported by the extensional content of the sequence. However, the tack of uniformly using S_1 is attractive on grounds of parsimony. We return to this question in the final section.

This section has introduced LFs for pictorial narratives that include introduction of discourse referents, equalities among discourse referents, an implicit conjunct L_1 that expresses extensional free perception, and an embedding operator S_1. Pictorial LFs are interpreted compositionally in dynamic possible worlds semantics. They can be mapped to a "surface" picture sequence by deleting all of the extra operators. Thus we assume a setup for grammar and semantic interpretation of pictorial narratives that in gross terms is the same as a typical setup in natural language semantics. It is worth noting that the interleaved operators are motivated by semantics—there is no other evidence for them. It follows that the syntax is relatively underdetermined. For instance, it would be possible to use the syntax $\Psi\ a\ L_1(q)$ in place of $\Psi\ a\ q\ L_1$, with adjustments in the semantics of L_1. Our version reflects the fact that the semantics is conjunctive.

[26]Maier and Bimpikou (2019) discuss examples in comics that are intensional, but where the viewpoint is neutral, rather than being the geometric viewpoint of the attitudinal subject. Section 4 gives similar examples from film and video series. This undercuts the argument from (25), because it shows that intensional semantics is consistent with a neutral projection viewpoint.

3. Event models of perception

This section reviews models that represent perceptual events and their epistemic consequences using alternative relations. We describe model structures that include events and alternatives to events, which together can encode things such as an individual seeing veridically that a man is approaching while taking off his hat, or an individual hallucinating that a girl is eating a rabbit sandwich. This lends substance to the analysis from Section 2, by filling in the model theory. And it prepares for a puzzle presented in Section 4.

Consider (26), which should be thought of as part of a narrative involving two characters Gable and Boyer playing cards with an ordinary deck.[27] Gable looks up, and sees Boyer holding up the ace of spades. The example is to be interpreted veridically, so that in a described situation, Boyer really is holding up an ace of spades. We treat Gable seeing Boyer holding up an ace of spades as an event e that takes place in the base world. Agents have epistemic access to their perceptual events, and so it is natural to posit that in epistemic alternatives for Gable, he participates in an event e' that is qualitatively the same as e. This is illustrated in (27) with a base world w_0, and alternative worlds w_1 and w_2 for Gable in w_0.

(26)

(27) *Sketch of model for Gable veridically seeing Boyer holding up an ace of spades, first version*

Base world w_0	Event e of Gable seeing Boyer holding up an ace of spades has just happened.
Alternative world w_1 for Gable in w_0	Event e' of Gable seeing Boyer holding up an ace of spades that is qualitatively equivalent to e has just happened.
Alternative world w_2 for Gable in w_0	Event e'' of Gable seeing Boyer holding up an ace of spades that is qualitatively equivalent to e and e' has just happened.

A neat way of encoding the assumption that e, e', and e'' are qualitatively the same is to assume that they are the same model theoretic object. On this account, there is a model-theoretic object e that has just happened in the base world w_0, and has also just happened in the alternative worlds w_1 and w_2. We say informally that e is an "event type" that can be "instantiated" by token events in different worlds, and by multiple token events in a single world. Suppose Gable has the exact same

[27] This is a constructed collage narrative, with the image of Clark Gable reflected, and a playing card superimposed on the image of Charles Boyer.

qualitative experience of seeing Boyer hold up an ace of spades twice in the same world-time line. This is represented by e having just happened in both w and w', where $w < w'$ (w' is a temporal extension of w). All of this leads to streamlining scenario (27) as in (28).

(28) *Sketch of model for Gable veridically seeing Boyer holding up an ace of spades, second version*

Base world w_0	Event e of Gable seeing Boyer holding up an ace of spades has just happened.
Alternative world w_1 for Gable in w_0	Event e has just happened.
Alternative world w_2 for Gable in w_0	Event e has just happened.

In theorization about events in linguistic semantics, events are standardly taken to be token events that cannot happen twice in the same world-time line. This is seen in axioms covering the temporal ordering of events, and in the fact that there is often posited a run-time *function* that maps events to their run-times.[28] This way of theorizing can be reconstructed in our setting by defining a token event to be an ordered pair of an event type and a world where that event has just been instantiated.[29] So while there is a single event type e, there can be worlds (at a time) w and w', where w is a prefix of w', e has just happened in w, and e has just happened in w'. In this configuration, e has happened twice in w', and there are distinct token events $\langle e, w\rangle$ and $\langle e, w'\rangle$ that have happened in w'. In the scenario (28) with world alternatives, while the model-theoretically "same" event e happens in w_0, w_1, and w_2, we can also refer to distinct token events $\langle e, w_0\rangle$, $\langle e, w_1\rangle$ and $\langle e, w_2\rangle$ happening in w_0, w_1, and w_2 respectively. Subsequently in this paper, we call an event type such as e simply an event, while saying "event token" when it is necessary to refer to world-bound individuals that happen at just one time.

Returning to the example, (29) is the LF according to Section 2 for a veridical interpretation of (26). The geometric point a sets up a discourse referent for the individual depicted in the first picture. The conjunct L_1 enforces coincidence between the viewpoint for the second picture and the agent's visual perspective, and imposes a presupposition that the agent just looked. This LF is verified by the model sketched in (28). Because of the alternatives, it describes Gable as picking up the information that Boyer has held up an ace of spades.

[28] See the axioms governing temporal precedence and temporal overlap among events in Kamp (1979).

[29] This is really a conjecture, because explaining it in detail requires recasting in these terms an explicit model-theoretic account of events in linguistic semantics, such an intensional version of the event models of Krifka (1989). This is a substantial enterprise. While the event models under discussion here differ from those typically used in linguistic semantics, technical aspects including logic and model theory are well developed. Literature on the topic is largely concerned with model theory and axiom systems for epistemic logic, and planning in artificial intelligence. The monographs Van Ditmarsch et al. (2007) and Reiter (2001) are good entry points. The motivation for using such models here is that they provide a substantive account of the epistemic consequences of looking events.

(29)  a  L_1

Example (30) is a different narrative, where in the second picture Boyer is holding up a moon tarot card instead of an ace of spades. The veridical interpretation has an LF isomorphic to (29), and has models like (31). This is just like (28), but with an event m of Gable veridically seeing Boyer holding up a moon tarot substituted for e, both in the base world and the alternatives.

(30)  

(31) *Sketch of model for Gable veridically seeing Boyer holding up a moon tarot card*

Base world u_0	Event m of Gable seeing Boyer holding up a moon tarot card has just happened.
Alternative world u_1 for Gable in u_0	Event m has just happened.
Alternative world u_2 for Gable in u_0	Event m has just happened.

What about Gable hallucinating? Consider the concrete scenario of Boyer holding up the ace of spades, but with Gable hallucinating that a moon tarot is held up, rather than perceiving the ace. In characterizing event models for this, we apply an idea from Baltag et al. (1998). Since Gable is hallucinating the moon tarot, his epistemic alternatives should have Gable facing a scene with a moon tarot, rather than a scene with an ace. In terms of event models, the alternative worlds should involve m having just happened, rather than e having just happened. Such models are schematized in (32). In the alternative worlds, the event m has just happened, meaning Gable sees a moon tarot in the alternative worlds.[30] The event in the base world is written h^m_a, where the superscript

[30] And indeed he does so veridically. This has consequences for epistemic propositions, for instance in world v_o, Gable

annotates that event m happens in alternative worlds, while the subscript indicates that Gable is facing a scene with an ace in the base world.

(32) *Sketch of model for Gable hallucinating that Boyer is holding up a moon tarot card, while Boyer actually holds up an ace of spades*

Base world v_0	Event h_a^m has just happened.
Alternative world v_1 for Gable in v_0	Event m has just happened.
Alternative world v_2 for Gable in v_0	Event m has just happened.

The discussion so far has schematically characterized models for veridical and hallucinating perception, in a way that refers to worlds and events occurring in these worlds. Where do assumptions such as e having just happened in any alternative to world w_0 of example (28) come from? It is not satisfactory to simply partially sketch intended models, as was done above. Research in dynamic epistemic logic proposes a systematic answer to this problem which is based on the technical idea of including alternative relations on events in the model construction, and deriving alternative relations on worlds from them (Baltag et al., 1998; Van Ditmarsch et al., 2007, 2015).[31] For the event e of Boyer holding up an ace and Gable veridically observing, the set of event-alternatives for Gable is stipulated to be the unit set of e. This correlates with the fact that in the schematic illustration (28), event e has just happened in every world alternative for Gable. (33) lists the event alternatives that are suitable for the scenario. To illustrate the formalism for multi-agent epistemic models, we list alternatives for Boyer as well. These alternatives encode Boyer knowing what card he is holding up.

(33)

event	alternatives for Gable	alternatives for Boyer	gloss
e	$\{e\}$	$\{e\}$	Boyer knowingly holding up an ace of spades, with Gable observing veridically
m	$\{m\}$	$\{m\}$	Boyer knowingly holding up a moon tarot card, with Gable observing veridically
h_a^m	$\{m\}$	$\{e\}$	Boyer knowingly holding up an ace of spades, with Gable hallucinating that Boyer is holding up a moon tarot card.

The next step is to construct possible worlds from events. A standard way of doing this relies on the idealization that events happen one at a time, so that a world at a time is a sequence of events. Where w is a world and e is an event, w is incremented to $w \diamond e$, using the partial operation called

believes that he believes Boyer just held up a moon tarot. This comes from m having just happened in all alternatives to worlds like v_1 and v_2.

[31] While the area of research is called dynamic epistemic logic, this refers to the dynamics of information of agents, not the dynamics of information in a discourse. The possible worlds models assumed in the current paper are static possible worlds models, where a world at a time is constructed as a sequence of events, and in the standard way epistemic alternative relations are relations between worlds. See Campbell and Rooth (2021) for a computational and algebraic development of this way of constructing multi-agent epistemic possible worlds models.

fusion product.[32] Incrementation is partial because events have preconditions. An elevator can go up a floor only if it is not already on the top floor. A light can go on only if it is off. An individual can veridically see a scene with an ace of spades only if she is facing such a scene. Preconditions of events are stative propositions, which are listed for each event in the construction of the model.

The discussion above of the veridical looking events e and m needs to be generalized, by making looking events and the information in them more specific. Pictures of Boyer holding up a card don't just have information about what card is held up. For instance, they have information about whether or not his eyes are closed, about what kind of neckwear he has on (scarf, necktie or bow tie) and about which hand he uses to hold up the card. To generalize the construction, we introduce primitive events of the form $l(x,q)$, where x is an agent and q is a visual scene. $l(x,q)$ is interpreted as the event of agent x looking veridically at a visual scene exactly like the visual scene q. Visual scenes are equated with pictures. The pre-condition for $l(x,q)$ happening in a world w is that w looks exactly like q from the visual-geometric perspective of agent x. Following the design sketched above, the set of event alternatives for $l(x,q)$ is the unit set $\{l(x,q)\}$. As explained in a moment, this entails that when agent x looks veridically at scene q, x picks up the information in q. In paraphrase, x picks up the information that he is in a world that looks like q from his geometric perspective. (34) records the assumptions about $l(x,q)$.[33]

(34) *Events of veridical looking*
Event $l(x,q)$ has the pre-condition that the world looks like q from the visual-geometric perspective of agent x, $\pi(w,V(w,x)) = q$. Event alternatives to $l(x,q)$ for agent x are the unit set of $l(x,q)$.

Let q_a be the ace picture in (29). If world w meets the pre-conditions of $l(x,q_a)$ (that is, if w looks like q_a from x's perspective), then w can be extended to $w \diamond l(x,q_a)$, which is a world where x has just looked veridically at the scene in front of him. In epistemic models, the information of agent x in a world is modeled by a set of world alternatives. Let U be the set of epistemic world alternatives for agent x in world w. In the approach from dynamic epistemic logic, the epistemic alternatives for x in the extended world $w \diamond l(x,q_a)$ is defined in terms of U and the set of event alternatives to $l(x,q_a)$. Specifically, alternatives for x in $w \diamond l(x,q_a)$ are worlds of the form $u \diamond e'$, where u is a world-alternative to w, and e' is an event-alternative to $l(x,q_a)$. In the case at hand, the sole option for e' is the event $l(x,q_a)$ itself, because the set of event alternatives to $l(x,q_a)$ is $\{l(x,q_a)\}$. Therefore any world alternative to $w \diamond l(x,q_a)$ is of the form $u \diamond l(x,q_a)$. Because of the precondition, in order for $u \diamond l(x,q_a)$ to be defined, agent x is facing the scene q_a in u. Consequently, in world $w \diamond l(x,q_a)$, agent x has the information that he is facing a scene exactly like q_a, because by definition this is true if every alternative to $w \diamond l(x,q_a)$ has x facing a scene exactly like q_a.[34]

In this way the event-alternative construction from dynamic epistemic logic captures the epistemic consequences of events. The method for extending event alternatives to epistemic world

[32]This setup can be identified with McCarthy's situation calculus (McCarthy, 1963; Reiter, 2001). In the situation calculus, a world w in our sense (i.e. a world at a time) is updated with an event e with the operation *do* to the world $do(e,w)$. w is called a situation and e is called an action. The term fusion product comes from guarded string models (Kozen, 2001).

[33]In Section 4, we argue that this adjustment goes too far, because the epistemic consequences of such events are too strong.

[34]Or technically, every alternative is of the form $u \diamond e'$, where u has x facing a scene exactly like q_a. In an idealized model, it can be stipulated that looking does not change the scene.

alternatives is stated in general form in (35). See Baltag et al. (1998) and Campbell and Rooth (2021) for mathematical developments of the construction.

(35) Let w be a world that meets the pre-conditions of event e, let U be the set of epistemic world alternatives for agent x in w, and let E be the set of event alternatives to e for x. Then $w \diamond e$ is a world, and the set of world alternatives for x in $w \diamond e$ is $\{u \diamond e' | w \in U \wedge e' \in E \wedge u \diamond e' \text{ is defined}\}$.

In the earlier discussion, h_a^m was described as an event of Gable facing a scene with an ace of spades in the base world, while hallucinating a moon tarot. This is generalized to $h(x,q,r)$, the event of agent x facing a scene like r in the base world, while hallucinating the view q. Like $l(x,r)$, this has the pre-condition that x is facing a scene like r. Its set of event alternatives is $\{l(x,q)\}$, the unit set of the event of x veridically seeing scene exactly like q. This has the consequence that in world $w \diamond h(x,q,r)$, where x has just hallucinated q, agent x has the information that he is facing a scene just like q. This follows from the same reasoning as for $w \diamond l(x,q)$.

We now restate the semantics for L_1 and S_1 from Section 2 in a way that is adapted to event models. For L_1 this is straightforward. The modified semantics (36) restates in event terms the presupposition that the agent picked out by index 1 has just looked. The assertion as before adds the information that the viewpoint v (this is the viewpoint for the second picture in free perception sequences) is the geometric viewpoint of the agent. A tuple of the form $\langle w \diamond l(g,q_a), v, \mathscr{O} \rangle$, where q_a is the ace picture, $\mathscr{O}[1] = g$, and v is g's visual-geometric viewpoint in world $w \diamond l(g,q_a)$, can satisfy the semantics coming from the different parts of the LF repeated in (37). Because of the general semantics of pictorial update, world $w \diamond l(g,q_a)$ looks like q_a from v. Because of the semantics of indexing and the assumption that $\mathscr{O}[1]$ is g (Gable), g is the individual depicted in the first picture. In combination, these give the information that the base world $w \diamond l(g,q_a)$ looks like q_a from g's visual-geometric perspective. Given the event alternatives for $l(g,q_a)$, any epistemic world alternative for g in $w \diamond l(g,q_a)$ is of the form $u \diamond l(g,q_a)$ This models Gable veridically seeing scene q_a, because in each of his epistemic alternatives, he faces a scene that looks like q_a, and has just looked at it. Thus in world $w \diamond l(g,q_a)$, he has picked up the information that he is facing the scene that looks like the scene he is actually facing, by looking at it.

(36) *Semantics of the extensional free perception constant L_1, event version, referring to a satisfying tuple $w, v, \mathscr{O}$*

Presupposition: $\mathscr{O}[1]$ is an agent with a visual system, and w is of the form $u \diamond e$, where e is a looking action by $\mathscr{O}[1]$.

Assertion: $v = V(w, \mathscr{O}[1])$ (v is the visual-geometric viewpoint in world w of agent $\mathscr{O}[1]$).

(37)  *a*  L_1

The LF for the moon card scenario with embedding is repeated in (38). Our approach to interpreting the embedding is to define a semantic relation $\widetilde{S}$, which is used in formulas like $\widetilde{S}(w,x,Q)$,

where w is a world, x is an agent, and Q is a viewpoint-centered proposition contributed as the semantic value of the embedded picture. The leading idea in defining $\widetilde{S}$ is to impose a condition of compatibility with the pictorial content in alternative worlds, and to preserve from the definition of L_1 the presupposition that the agent has just looked. Together with reference to the geometric viewpoint of the agent in alternative worlds, this gives the embedding predicate S_1 a semantics that is specifically connected to looking.

(38)  a S_1 

(39) is the semantics for $\widetilde{S}(w,x,Q)$ that is appropriate to current assumptions. The presupposition is that agent x has just looked in the base world w, by participating in a looking event e. The statement of the assertion takes advantage of the fact that alternatives to world $w' \diamond e$, where e is an event, are of the form $u' \diamond e'$, where e' is an alternative to e. It is this event e' that should be required to have preconditions that are consistent with the pictorial content Q. In the statement of the assertion, the alternative world u is decomposed as $u' \diamond e'$, and u' is constrained to satisfy the pictorial content Q. Because pictorial contents are viewpoint-centered, expressing this involves mapping agent x to a viewpoint, using the operator V that maps a world and an agent to the geometric viewpoint of the agent in the world. The effect of the definition is to require that in any alternative world, the agent is involved in an event of looking at a scene that looks like the embedded picture.[35] (40) is the inductive clause for a narrative incremented with a picture embedded under S_1. It uses the relation $\widetilde{S}$ to interpret the syntax $S_1(q)$. In the narrative, q is a syntactic picture. The corresponding pictorial content is obtained using the projection operator π.

(39) *Semantics of pictorial embedding relation* $\widetilde{S}$

Let w be a world, x an agent, and Q a relation between worlds and viewpoints, and let R be the epistemic alternative relation. Then $\widetilde{S}(w,x,Q)$ has the following semantics.

Presupposition: w is of the form $w' \diamond e$, where e is a looking action with agent x.

Assertion: $\forall u[R(w,x,u) \rightarrow \exists u' \exists e'[u = u' \diamond e' \wedge Q(u', V(u',x))]$

(40) *Incrementing with* $S_1(q)$*, event version*

$$\frac{w,v,\mathscr{O} \models \psi \quad \widetilde{S}(w,\mathscr{O}[1],\lambda w'v'.\pi(w',v') = q)}{w,v,\mathscr{O} \models \psi\, S_1(q)}$$

Using this definition and the event model construction, we argue that a world of the form $w \diamond h(g,q_m,r)$ satisfies the narrative (38). The event $h(g,q_m,r)$ is glossed as Gable hallucinating the moon card scene, while facing a scene like r in the base world. Definition (40) is applied with $w \diamond h(g,q_m,r)$ as the world and with the assumption that $\mathscr{O}[1] = g$, the individual Gable. The

[35] e' is a looking event because it is an alternative to the looking event e. This comes from the model construction rather than the definition.

presupposition in (39) is satisfied because $h(g,q_m,r)$ is a looking action by g, though a hallucinating one. The assertive condition involving $\widetilde{S}$ in (39) is $\widetilde{S}(w\diamond h(g,q_m,r),g,\lambda wv.\pi(w,v)=q_m)$ in the case at hand. In (41), this is verified with a chain of equivalences. The first line develops the condition as defined in (39). The second equivalence applies beta reduction. The third one applies definition (35) to $R(w\diamond h(g,q_m,r),g,u)$, reducing it to a condition on world alternatives to w, and event alternatives to $h(g,q_m,r),g,u)$. R^E is the three-place event alternative relation, relating an event, an individual, and an alternative event. The fourth equivalence exploits the fact that the sole event alternative to $h(g,q_m,r),g,u)$ for g is $l(g,q_m)$, the event of Gable veridically looking at the moon card scene. The last equivalence uses the fact that for $u'\diamond l(g,q_m)$ to be defined, u' has to satisfy the precondition of $l(g,q_m)$. This precondition as defined in (34) is $\pi(u',V(u',g))=q_m$, or in words world u' looks like the moon card scene from Gable's geometric perspective in u'. In the last line, reference to alternative relations has been eliminated, and this line is a theorem of predicate logic. The derivation is possible because world-alternative relations are defined in terms of event-alternative relations, so reasoning about epistemic world alternatives can be reduced to reasoning about events. And since alternative relations for events are defined in terms of explicit sets of alternatives for each event, reference to alternative relations on events can be eliminated as well.

(41)
$$\begin{aligned}
&\widetilde{S}(w\diamond h(g,q_m,r),g,\lambda wv.\pi(w,v)=q_m)\\
&\equiv \forall u[R(w\diamond h(g,q_m,r),g,u)\rightarrow \exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ [\lambda wv.\pi(w,v)=q_m](u',V(u',g))\end{array}\right]\\
&\equiv \forall u[R(w\diamond h(g,q_m,r),g,u)\rightarrow \exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ \pi(u',V(u',g))=q_m\end{array}\right]\\
&\equiv \forall u\left[\exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ R(w',g,u')\wedge\\ R^E(h(g,q_m,r),g,e')\end{array}\right]\rightarrow \exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ \pi(u',V(u',g))=q_m\end{array}\right]\right]\\
&\equiv \forall u\left[\exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ R(w',g,u')\wedge\\ e'=l(g,q_m)\end{array}\right]\rightarrow \exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ \pi(u',V(u',g))=q_m\end{array}\right]\right]\\
&\equiv \forall u\left[\exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ R(w',g,u')\wedge\\ e'=l(g,q_m)\\ \pi(u',V(u',g))=q_m\end{array}\right]\rightarrow \exists u'\exists e'\left[\begin{array}{l} u=u'\diamond e'\wedge\\ \pi(u',V(u',g))=q_m\end{array}\right]\right]
\end{aligned}$$

This reasoning shows explicitly that world $w\diamond h(g,q_m,r)$ satisfies the narrative (38). This supports the event-sequence model as a workable model of the epistemic consequences of hallucinating events. And it tends to support the adequacy of the embedding LF with S_1, and the associated semantics in its event form.

This completes the development of a basic event model for perceptual events, and a semantics for L_1 and S_1 in such models. The key idea is to model the epistemic consequences of events by stipulating alternatives for events in the model construction, and to use events of veridical looking as the event alternatives for events of hallucination. The semantics for L_1 includes a presupposition that the agent picked out by discourse referent 1 has just looked, and a constraint that the viewpoint for the second picture is the geometric viewpoint of that agent. The semantics for the embedding predicate S_1 also has a presupposition that agent 1 has just looked, and in the assertion quantifies

universally over world alternatives.

4. Over-informative complements

The analysis so far has used Hintikka semantics for intensional embedding, involving universal quantification over alternatives, with adaptations for the pictorial setting. In the semantics from the end of Section 3, in a verifying scenario for an intensional LF, each alternative for the agent is required to have the counterpart of the agent facing a scene exactly like the embedded picture. This is in one respect problematic. When human agents look at a scene, they typically do not pick up all of the information in the scene. Suppose Keisha is sitting at the table in her dining room, reading about the brood of locusts which are emerging in her area. She looks up from her laptop at the glass door facing her patio, where hundreds of locusts are on the outside surface of the glass and beyond it. In a scenario of veridical perception, in Keisha's actual situation, there is some specific number of locusts visible from her vantage point, say two hundred and eleven. But she would not and normally could not pick up in a glance the information that there are two hundred and eleven locusts visible from her vantage point. Rather, she would pick up information along the lines there being a lot of locusts, together with more specific information about some of them. Suppose that, rather than seeing locusts veridically, Keisha imagined or hallucinated them. Here too Keisha can imagine a bunch of locusts outside the glass door, without imagining that there is any specific number.

The above observations are part of our ordinary experience. They are also reflected in language data. Literature on perception-verb complementation points out that whether perceptual complementation has strong entailments correlates with the complement type (Barwise, 1981; Higginbotham, 1983). Example (42) has the strong entailment that Keisha gets information that entails the complement sentence, including the numerical information. Because picking up the numerical information is implausible, the example is odd.[36] Example (43) with a base-form tenseless complement is more neutral about what information is picked up.[37] It does not entail that Keisha picks up the numerical information.

(42) *Tensed complement*
Keisha saw that two hundred and eleven locusts were crawling on the outside surface of the glass door.

(43) *Tenseless complement*
Keisha saw two hundred and eleven locusts crawl on the outside surface of the glass door.

The same distinction applies in varieties of implicit embedding. (44) is an example of free

[36] In principle it is necessary to guard against wide scope interpretation of the DP [two hundred and eleven locusts], producing a reading that is paraphrased by (i), which doesn't attribute numerical information to Keisha. Intuitively we don't get this reading for (42). Readers who do should consult their intuitions about (ii) in place of (42), where there-insertion rules out a wide-scope LF (Heim, 1987).

(i) Two hundred and eleven locusts are such that Keisha saw them crawling on the outside surface of the glass door.

(ii) Keisha saw that there were two hundred and eleven locusts on the outside surface of the glass door.

[37] An ING-complement is also possible, but here the possibility of a DP complement with a post-modifier, rather than a clausal complement, must be controlled for.

(i) Keisha saw two hundred and eleven locusts crawling on the outside surface of the glass door.

indirect discourse (*FID*).[38] This is an implicit embedding construction that has a semantics similar to tensed clausal embedding (Banfield, 1982; Schlenker, 2004; Sharvit, 2008; Eckardt, 2014).[39]

(44) *Free indirect discourse*
Keisha looked up with dismay. Tomorrow was Christmas, and two hundred and eleven locusts were crawling on the outside surface of the glass door. Cripes, Santa Claus might be scared off, and not come in to leave the presents.

Free indirect discourse has certain identifying features and interpretive properties, several of which are seen in (44). The normally deictic *tomorrow* is understood to pick out the day after Keisha's looking time, and thus is "shifted" by the implicit embedding. There is an emotive interjection *cripes*, which is understood to describe Keisha's reaction of dismay. And in interpretation, the modal *might* seems to describe possibility according to Keisha. Literature on FID argues that the features correlate and thus identify the construction. FID is strong in the information it attributes to the attitude holder. Intuitively it seems clear that FID example (44) entails that Keisha picks up the numerical information, just as in example (42) with a tensed complement. Consequently the FID example seems odd. (45) is a different construction, which we call a free perception report, and identify with viewpoint shifting in the sense of Hinterwimmer (2017) and protagonist projection in the sense of Abrusán (2021). These examples do not entail that Keisha picks up the numerical information (or in general, all of the information in the scene before her), and are thus possible descriptions of events where Keisha simply sees a profusion of locusts.[40]

(45) *Linguistic free perception report*

a. Keisha looked up from her laptop. Two hundred and eleven locusts were crawling on the outside surface of the glass door.

b. When Keisha looked up from her laptop, two hundred and eleven locusts were crawling on the outside surface of the glass door.

To make claims about these data precise, consider a specific world $w \diamond e$, where e is an event of Keisha looking at a glass door with two hundred and eleven locusts outside the glass, with Keisha picking up the information that there are a bunch of locusts, but not the numerical information or detailed geometric information about all of the locusts. Let φ be the clause given in (46), which is not entailed by Keisha's information in $w \diamond e$. Sentences with tenseless embedding of φ and sequences with free-perception embedding of φ are truthful descriptions of $w \diamond e$ (see (43), (45a),

[38](44) is a version of the Christmas example from Zimmermann (2008).

[39]Eckardt, rather that postulating implicit embedding, uses non-deterministic semantic interpretation (Eckardt 2014, p. 75). In the framework used in this paper, the LFs that are compositionally interpreted are disambiguated, so there is no non-deterministic interpretation. We are unsure whether there is any substantial issue between the two styles of analysis. Certainly, readers have the option of interpreting a given passage as FID or not, unless there are disambiguating features. But this applies to any kind of ambiguity. It is relevant that Eckardt in part motivates her analysis with the claim that there are restrictions on multiple embedding of FID, although she cites an example of it that is characterized as being of a restricted form (Eckardt 2014, p. 57-58). Relatedly, Cumming et al. (2021) argue that there are restrictions on multiple embedding of point-of-view shots in film. In the analysis developed here, if embedding is a root phenomenon for some operators, that can be stated in the syntax. Constraints can also come from the model. It is impossible for ordinary humans to veridically see what another human is hallucinating.

[40]By Hinterwimmer's criteria, (45b) is viewpoint shifting and not FID, because it has the implicit embedding clause in construction with a when-clause, something that is not possible for FID. (45a) is in principle ambiguous, but it has a non-FID-reading of viewpoint shifting or free perception, as evidenced by the possibility of a weaker semantics.

and (45b)). Sentences with tensed embedding of φ and sequences with FID-embedding of φ are not truthful descriptions of $w \diamond e$, because they attribute too much information to the agent (see (42), (44)). This is explained if Hintikka semantics is used for tensed complements and FID, while a weaker semantics is used for the other constructions.

(46)

World $w \diamond e$	Keisha has just looked at a glass door with two hundred and eleven locusts without gaining numerical information or detailed geometric information.
Clause φ	Two hundred and eleven locusts crawl on the outside surface of the glass door

In this pattern, pictorial free perception patterns with tenseless embedding and linguistic free perception, and not with tensed embedding and FID. (47) is the pictorial version of the example. Intuitively it is neutral about how much information Keisha (the individual depicted in the first panel) picks up. It seems to be simply implied that she looked at the scene in some normal way. On an extensional interpretation/scenario, this is consistent with Sections 2 and 3, because an extensional LF entails that Keisha looks in the base world, and that she is facing a scene like the second picture in the base world. But it does not entail that she is facing a scene like the second picture in every alternative world.

(47)

Artist Milka Green

So much so good. The problem comes with the intensional LF and the semantics for it in Section 3. According to that semantics, when the intensional LF shown in (48) is true in a world $w \diamond e$, e is a looking action by the agent, and every epistemic alternative for the agent is of the form $u' \diamond e'$, where in world u' the agent is facing a scene exactly like the embedded picture.

(48) *Pictorial intensional free perception report*

$a \quad S_1$

Consider a specific world $w \diamond e$ which is a scenario of ordinary hallucination. By ordinary hallucination, we mean an event where the agent picks up some non-actual information and does not reject it, and where the experience is introspectively the same for the agent as an experience of veridical perception. On the semantics and model theory from Sections 2 and 3, if the LF (48) is true in $w \diamond e$, then every alternative for Keisha is of the form $u' \diamond e'$, where in u', Keisha is facing a scene exactly like the embedded picture. In epistemic semantics, this models the agent *having* picked up all of the information in the second picture. This means that, given the semantics from Sections 2 and 3, the sequence describes Keisha as gaining much more information in the event of hallucinating than she gains in events of veridical perception. This is an odd and unacceptable consequence: if Keisha has good powers of epistemic introspection, she could figure out that she must be hallucinating, by observing that she gained more information than she could gain with veridical looking. This should not be true of an event of ordinary hallucination, which is introspectively equivalent for the agent to an event of veridical looking.

Another way of posing the problem is that the analysis from Sections 2 and 3 claims that pictorial free perception on an intensional construal has a semantics like clausal embedding and FID, assuming those are analyzed with Hintikka semantics. But empirically, pictorial free perception, even on an intensional construal, is parallel to linguistic tenseless embedding and linguistic free perception, with a weaker semantics.

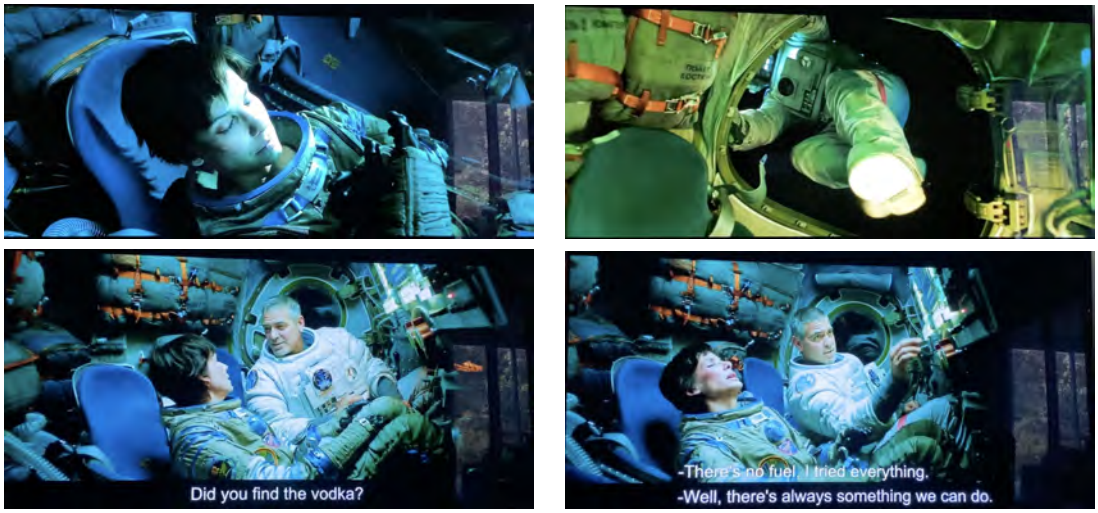
We call this the problem of *over-informative pictorial complements*. In our analysis and argument, the problem comes up for intensional LFs, and not extensional LFs. In the latter, the second picture describes the appearance of the base world from the geometric perspective of the agent. What alternative worlds are required to look like is determined by the event-alternative relation for the looking event, not the free-perception semantics.

The problem of over-informative complements comes up also for intensional passages which do not assume the visual perspective of a character.[41] In film, these are intensional passages that are not composed of point-of-view shots, in that the shots do not assume the visual perspective of a character. And it comes up, in addition to passages with an interpretation of hallucination,

[41]Bimpikou (2018) and Maier and Bimpikou (2019) are studies of neutral-viewpoint intensional panels in comics, looking at attitudes including imagination and giving many examples. They call neutral-viewpoint intensional panels "blended" panels, echoing one of the accounts in Bimpikou (2018), where information from the base world and information from world alternatives is blended geometrically in one panel. We assume that embedding phenomena are the same or similar in comics and film. On the geometric blending approach, see also Abusch and Rooth (2022).

for passages with an interpretation of imagining and passages with an interpretation of memory. Here we give a handful of examples from film and television series.[42] In each case, the point is that the semantic content of the passage is very strong compared to the attitudinal information of the agent in plausible verifying worlds. In the film *Gravity*, a scientist-astronaut Stone is on a space mission, and after several catastrophic accidents, including events that apparently cause the death of her companion, the seasoned astronaut Kowalski, she finds herself seemingly without fuel and with limited oxygen in a Soyuz capsule (Cuarón, 2013). She gives up on survival, turns off the oxygen supply, and appears to begin to lose consciousness. Thereupon, in a dream or reverie of Stone's, Kowalski is shown entering the capsule from the exterior, speaking to Stone in a way that bucks up her spirit, and suggesting a way for her to save herself and return to Earth. The passage is shot from a neutral perspective, showing Kowalski, Stone, and the interior of the capsule. The passage has rich visual and audio detail. The characters assume particular postures at particular times. The interior of the capsule is filled with instrumentation, dynamic lights and displays, Russian writing in the Cyrillic alphabet, hundreds of detailed objects such as containers, straps, wavy fabric, switches and dials, and objects floating about in the weightless environment. Indicator lights go on and off at particular times. In Kowalski's spoken language, token words, syllables and phones have particlar durations, his voice has a particular pitch at a particular times, and his spoken utterances have particular acoustic waveforms. The passage is three and a half minutes long.

(49) From *Gravity*, Cuarón (2013).
Sandra Bullock (Stone) and George Clooney (Kowalski).



In Episode 8, Season 3 of the television series *Poldark*, the protagonist has met an old flame in the village graveyard, and then kissed her in the village church (Barry, 1975). In a subsequent

[42]The problem of over-informative complements is more compelling for film than for comics, for several reasons. Intensional constructions are common in film, and there is an excellent basis of empirical observation and theory (e.g. Turim 1989). Second, film shots have far more information than comic panels, and so what has to be entailed in the analysis from Section 3 is stronger. Third, the program of possible worlds semantics for pictorial narratives based on geometric projection is more easily applied in film, because normal film shots are geometrically accurate. The question of how to deal with stylized or otherwise non-realistic pictures has not been faced squarely in literature to date on possible worlds semantics for comics.

passage, Poldark imagines confessing the incident to his wife. (50) shows one frame. The passage is shot from a neutral perspective and is seventy-five seconds long. It is visually and acoustically detailed. Poldark's body and face assume particular shapes at particular times. His imagined speech has a particular detailed waveform that is conveyed in the audio for the passage. The shot shows features in the background that are out of view of the depicted character.

(50) From *Poldark*, Barry (1975). Robin Ellis (Poldark)
Poldark: I met Elizabeth. For the first time in years, we talked.

In the film *A Beautiful Mind*, the character Nash, in delusional passages, imagines interactions with characters including a Department of Defense official, a friend met in grad school, and the friend's niece (Howard, 2001). Nash is shown speaking to the imagined characters. These passages are visually, temporally, and acoustically detailed.

The silent film *The Woman on Trial* is principally concerned with a trial and testimony in it (Stiller, 1927). A woman is charged with conspiring in the murder of her lover. In trial testimony, she describes an alternative sequence of events involving revenge perpetrated by her husband. This testimony is shown in flashback, and it is inferred that the content does not describe the base world of the narrative.[43]

In the style of analysis from Section 2, influenced by linguistic theory, the above passages can be argued to have a logical form of embedding.[44] If the passages were not embedded, their content would survive as an entailment of the narrative as a whole. And the Poldark episode, for instance, is not understood as entailing that Poldark did confess to his wife. So, we analyze the Poldark confession passage as having a logical form $Im_1(\varphi)$, where φ is the confession shot, *Im* is an imagination attitude, and 1 is the index for Poldark. Similarly for the other passages. For the passage from *Gravity*, and the passages from *A Beautiful Mind* we postulate an implicit embedding predicate *Ha*, for hallucination. This is similar to the embedding predicate *S* from Section 3, but is distinct, because the shots and passages in these cases are not point-of-view shots.

The passages from *The Woman on Trial* are open to a different interpretations, all of them involving embedding in the framework assumed here. Turim (1989) argued that flashbacks in early film often have an interpretation of memory. The passages could be construed as embedded under a hidden memory operator, with LFs like $Me_1(\varphi)$.[45]

[43]See the discussion of *The Woman on Trial* in Turim (1989), p. 53 in the section *Trial Testimony Flashbacks*, and in the section *Flashbacks Rendering Verbal Narration Visual*, p. 49.

[44]The terminology of embedding is also current in narrative theory and film theory (Nelles, 2002, 2010).

[45]But alternatively, embedded shots in trial scenes (especially in silent films) can be seen as substituting for sentences spoken by the witnesses. "It is possible to see the use of the flashback in the late silent period as part of the compensation for the lack of sound and as an anticipation of the use of sound. By substituting flashback for a story told

In all of these examples, the embedded passage that is part of the film has very strong geometric, acoustic, and temporal content. Is it plausible to assume models where the attitudinal information of characters in described situations is this strong? Letting *Gravity* be the film about Stone and Kowalski as a whole, its semantic value $[\![\textit{Gravity}]\!]$ is a set of worlds (in the sense of worlds at a time, or alternatively entire world-timelines), each of which includes events of Stone imagining Kowalski entering the capsule. Let w be one of these worlds, and let i_1^w ... i_n^w be the events of Stone imagining in w. How specific is the information content of these events of hallucinating? In w, did Stone hallucinate (or dream) the particular time course of the many blinking lights and of the images on the dynamic displays? Did she hallucinate the particular folds of the fabric and configuration of the straps behind the head and out of view of Stone's counterparts? Did she hallucinate the particular acoustic waveforms of Kowalski's utterances? Our short answer to these questions is "obviously no".[46]

The situation is if anything worse for attitudes of imagining. In the embedded shot from *Poldark*, Poldark is described as imagining a confession to his wife. The information content of events of imagining can be more or less strong. It is plausible that in described situations, Poldark went over in his mind the sentences that he is shown uttering. In the possible-worlds model, this is modeled by his imagination events having alternatives where he speaks the same words. But there is no motivation at all for maintaining that the imagination events have geometrically detailed information about the background of the part of the room that is out of view of counterparts of Poldark (such as the configuration of the curtains) or acoustically detailed information about his depicted utterances, such as their detailed acoustic waveforms.

There are a couple of issues that complicate this argument in interesting ways, but hardly weaken it. There is radical variation among people in the richness of experiences of imagining. Some people have intensely detailed visions, a condition called hyperphantasia. The neuroscientist Joel Pearson, as quoted in Zimmer (2021), describes the experience of such people as "like having a very vivid dream and not being sure if it is real or not," and "people watch a movie, and

verbally, the silent cinema could express complicated dialogue or monologue scenes without resorting to a plethora of wordy titles" (Turim, 1989). This calls for a different LF, with an embedding predicate like *say*, which however still has a pictorial complement. This would be a clear case of the information in the base world (what the witness said) not entailing the geometric and temporal content of the film shot.

[46] An L&P reviewer brought up phenomena of non-diegetic sound and absence of blinking in film as counterpoints to over-informative embedded shots. Bordwell (2012) points out that people are seldom depicted as blinking in film, while in reality they blink ten to twenty-five times per minute. Consider a two-minite shot of a character's face, who is never shown blinking. Stipulate that it is physiologically impossible for a person with open eyes not to blink over a two-minute span. Then on the face of it, the shot is non-naturalistic, because described situations for the shot are inconsistent with human physiology. There is the option of solving the non-naturalism in the syntax-semantics interface, by inserting blinks randomly to obtain naturalistic described situations. Or one could accept that described situations are non-naturalistic. Non-diegetic sound (such as film score) is by definition not mapped from the described situation. But in principle, one could accept that described situations for films with scores include orchestral sounds of no physical origin in the described world, so that described situations are physically non-naturalistic. Or one can solve the non-naturalism in the syntax-semantics interface, by mixing a score into the soundtrack of the film, external to semantic interpretation. Here considerations of normality might be involved in reducing ambiguity. The analysis in this paper addresses the problem of over-informative complements in the syntax-semantics interface. In parallel to that, non-diegetic sound should be addressed in the syntax-semantics interface. For the blinking problem, it isn't obvious to us whether the "right" solution is in the syntax-semantics interface, or whether non-naturalistic models should be accepted. An issue here is whether the aesthetic consequences of absence of blinking are to be analyzed in terms of semantic values. This would be impossible if blinks have been inserted in the semantic values of shots without blinks.

then they can watch it again in their mind, and it's indistinguishable." According to Zeman et al. (2020), hyperphantasics comprise around 2.6 percent of the population.[47] One could try to claim that visual narratives present all characters as being hyperphantastics, so that when they imagine they have experiences (or believe they have experiences) that are just as informationally strong as their experiences of ordinary looking. But this would not help, because the point of the argument above is that an analysis is unacceptable that predicts that the information gained in events of hallucinating and imagining is *far stronger* than the information gained in events of veridical looking.

A related point is that people over-estimate the information content of events of veridical looking. When we look at our environments, our eyes dart about between fixations, and we get detailed information only about visual regions that are fixated. Apart from this, we may fail to "attend" to objects in our visual fields, such as a gorilla walking through the scene (Simons and Chabris, 1999; Chabris and Simons, 2010).[48] Observers may fail to notice that two individuals in their field of view exchange heads (Grimes, 1996), or that the individual they are talking to on the street is replaced by another person in an interval when their view is blocked (Simons and Levin, 1998). This phenomenon is known as "perceptual blindness" or "inattentional blindness". The fact that people find these experimental results surprising indicates that they over-estimate the information content of their ordinary visual experience. But it is also not part of our ordinary conceptions that we pick up *all* of the information in our visual environments.

This brings up the issue whether, in formulating a semantics for pictorial free perception, we should employ models that reflect scientific understanding of human perception and cognition, or models that correspond to some kind of folk theory.[49] If models reflect scientific understanding, then given perceptual blindness, the information of viewers is certainly far weaker than the information in the visual scenes they are facing. According to the metaphysical stance of Lewis (1986), possible worlds as employed in philosophical theories have the same character as the world we occupy. And we take it that most practicioners of linguistic semantics adopt a metaphysical stance where possible worlds, including those that figure in epistemic semantics, are realistic in this sense. However, we think there is no definitive answer to the question whether scientifically valid models or models reflecting ordinary understanding (or a metaphysics implicit in language) should be employed, and so both kinds of models should be investigated.[50] For the argument being made here it does not matter, in that also according to our common understanding, people do not pick up all of the information in their visual environments when they look, and their attitudinal information when they hallucinate is at most as strong as when they look veridically.[51]

Our strategy for solving the problem is a three-part one. First, a model construction is developed

[47]The opposite condition is known as aphantasia (Zeman et al., 2016).

[48]Participants view a video clip of six people, three in white shirts and three in black shirts, passing basketballs to each other. They are asked to keep a silent count of the number of times an individual in a white shirt makes a pass. In the middle of the clip, an individual in a gorilla suit walks into the center of the scene, thumps their chest, and leaves the scene. Half of the participants who successfully counted passes did not notice the gorilla, as measured in a post-test.

[49]Thanks to Emar Maier for raising this issue.

[50]The reason is that the gap between scientifically naturalistic models and typical models from natural language semantics is so wide. For instance, macroscopic individuals are built into the model structure in the latter.

[51]The only thing that would help with the argument being made here is to postulate that agents as presented in visual narratives do pick up as much information as is recorded in pictures and film shots, and that when they hallucinate, imagine, or remember, their experience is informationally as strong as when they look veridically. See Section 6 for some discussion of this option.

that takes the relatively weak epistemic consequences of looking into account, by holding constant across epistemic alternatives only those features that are attended to. Then a normality definition is developed, which defines the normal epistemic consequences of looking at a given scene. Finally, the normality condition is incorporated in the semantics for intensional free perception. This is accompanied by weakening quantificational force in the semantics of intensional embedding.

We motivate the revised analysis by reconsidering the semantics and model construction for sequences that are construed veridically. In the sequence repeated in (51), the syntax-semantics interface enforces that in a described situation, Gable is facing a scene just like the ace picture when he looks. This much is fine. But it should not follow that Gable picks up all of the information in the ace picture. If he is a human with normal perceptual procedures, or a human who behaves in accordance with our ordinary conception of things, he would normally *not* pick up all of the information in the scene he is facing.

(51)  a  L_1

A model is required where there are events of Gable looking veridically while facing a scene like the ace picture that are weaker in their epistemic consequences than the event $l(g, q_a)$ of Gable looking at the ace scene and picking up all the information in it. Since Gable is playing cards, it might be that in any normal course of events where Boyer holds up an ace of spades, Gable focalizes that card and attends to it, and picks up the information that it is an ace of spades.[52] But in normal courses of events, Gable might not pick up information about Boyer's neckwear, or whether Boyer's left eye is partially closed.

Building possibilities like these into the event model is as complex and varied as the possibilities that we would like to allow for. Here, to illustrate ideas, we consider just three dimensions of variation in the scene: (i) the card that Boyer is holding up; (ii) the nature of Boyer's neckwear, e.g. necktie vs scarf; (iii) the configuration of Boyer's eyes, e.g. open or closed. Let d_{xyz} be the event of Gable looking at a scene with Boyer holding up card x, while wearing neckwear y, and with eyes in configuration z. Looking events where Gable attends to a given feature should be distinguished from ones where he ignores that feature. This is recorded in events by bold-facing the features that Gable attends to. For instance, $d_{\boldsymbol{a}\mathbf{n}o}$ is the event of Gable looking at a Boyer-like man holding up an ace of spades (letter $\boldsymbol{a}$ in the subscript) who is wearing a necktie (letter **n** in the subscript) with open eyes (letter o in the subscript), while attending to the card (boldfacing of first letter in the subscript) and to the neckwear (boldfacing of second letter in the subscript) but not the configuration of the eyes. Event $d_{\boldsymbol{a}no}$ is the event of Gable looking at a Boyer-like man holding up an ace of spades who is wearing a necktie and has open eyes, while attending to the card (boldfacing of first letter in the subscript) but not the neckwear or the configuration of the eyes.

[52] An L&P reviewer brought up the fact that zoom, dolly movement, and depth of field are used in film to convey focalization. While these reduce or modify information in the direction of information picked up by the depicted agent, such shots are still over-informative. Nevertheless there is an implication that the agent focalizes and attends to *some* parts of the zoomed scene, or some part of the scene that is in focus in the case of shots with limited depth of field. This isn't treated by the formalization in this paper.

event	precondition	alternatives	gloss
$d_{\boldsymbol{a}no}$	Gable faces ace, necktie, and open eyes	$\{d_{\boldsymbol{a}no}, d_{\boldsymbol{a}nc}$ $d_{\boldsymbol{a}so}, d_{\boldsymbol{a}sc}\}$	Gable looking veridically at the ace scene while attending only to the card
$d_{\boldsymbol{m}no}$	Gable faces moon, necktie, and open eyes	$\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}nc}$ $d_{\boldsymbol{m}so}, d_{\boldsymbol{m}sc}\}$	Gable looking veridically at the moon scene while attending only to the card
$d_{a\boldsymbol{n}o}$	Gable faces ace, necktie, and open eyes	$\{d_{a\boldsymbol{n}o}, d_{a\boldsymbol{n}c}$ $d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}\}$	Gable looking veridically at the ace scene while attending only to the neckwear
$d_{m\boldsymbol{n}o}$	Gable faces moon, necktie, and open eyes	$\{d_{a\boldsymbol{n}o}, d_{a\boldsymbol{n}c}$ $d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}\}$	Gable looking veridically at the moon scene while attending only to the neckwear
$d_{\boldsymbol{an}o}$	Gable faces ace, necktie, and open eyes	$\{d_{\boldsymbol{an}o}, d_{\boldsymbol{an}c}\}$	Gable looking veridically at the ace scene while attending only to the card and neckwear
$d_{\boldsymbol{mn}o}$	Gable faces moon, necktie, and open eyes	$\{d_{\boldsymbol{mn}o}, d_{\boldsymbol{mn}c}\}$	Gable looking veridically at the moon scene while attending only to the card and neckwear

Figure 2: Examples of events in a model of veridical looking and attention. Features that are attended to are boldfaced, and these features remain constant in event alternatives.

In the construction from Section 3, the set of event of alternatives to $l(g,q)$ was the unit set $\{l(g,q)\}$, and this modeled the agent Gable picking up all the information in the scene he is facing. How should this be modified when the event is extended by adding feature values for the card, the neckwear, and the eye configuration, and for Gable's attentiveness to the three features? The principle for this is that features that the agent attends to remain constant in alternatives, while other features can vary. Assuming that there are two eye states o (open) and c (closed), the set of event alternatives to $d_{\boldsymbol{an}o}$ for Gable is $\{d_{\boldsymbol{an}o}, d_{\boldsymbol{an}c}\}$, where there is variation in the eye-position features. Assuming there are two kinds of neckwear n (necktie) and s (scarf), the set of event alternatives to $d_{\boldsymbol{a}no}$ for Gable is $\{d_{\boldsymbol{a}no}, d_{\boldsymbol{a}nc}, d_{\boldsymbol{a}so}, d_{\boldsymbol{a}sc}\}$, where there is variation in the neckwear and eye features, but not the card feature.

Events like this are a model of the informational weakness of looking events in our ordinary conception of things. Or alternatively they model informational weakness of events of looking according to the actual nature of perception, as it is scientifically understood. When such events occur in a base world, agents pick up some information from the scene they are facing, but not all. In the card game, Boyer may hold up the ace, and Gable attends to that, but not to Boyer's neckwear or to the configuration of Boyer's eyes. In this way, Gable is "blind" to the necktie and the open eyes, like a participant in the Simons and Chabris experiment who is blind to the appearance of a gorilla.

In event models with attention features as defined here, pre-conditions of events are independent of the attention features. For instance, it is a pre-condition of $d_{\boldsymbol{a}no}$ that the individual Gable views is holding up an ace, wearing a necktie, and has open eyes, even though Gable does not attend to the neckwear or the eyes. Consider event $d_{\boldsymbol{a}no}$ transpiring in the base world, glossed as Gable looking at a scene with an ace, a necktie, and open eyes, while attending to the card but not the neckwear or the eyes. Weakened epistemic consequences are modeled by Gable having world alternatives that finish with any of the events $d_{\boldsymbol{a}no}$, $d_{\boldsymbol{a}nc}$, $d_{\boldsymbol{a}so}$, or $d_{\boldsymbol{a}sc}$, where Boyer's neckwear and Boyer's eye state vary across alternatives. As before, world alternatives are induced from event alternatives. Figure

2 lists some events of veridical looking in the attention model, and their alternatives for Gable.

The narrative snippet (51) is naturally read as entailing that Gable gains the information that an ace is held up, while remaining neutral about what other information in the second panel Gable picks up. We suggest this comes from the kind of normality implicature that is typical in the interpretation of narratives. The logic of the interpretation is roughly "Gable was facing a scene exactly like the second panel, and took a perceptual action that he could normally take while facing a scene like that." To this we add the assumption that in a situation where Gable and Boyer are playing cards and Boyer holds up a card, Gable would normally attend to it. Which events are normal ways of looking varies from situation to situation. In our scenario, since a card game is going on, if Boyer holds up a card, Gable would normally attend to it. But if Gable is active in an expert panel on neckwear, he would normally attend to the neckwear. Further, what is attended to in a scene depends on the visual salience of objects. We will not build this into our toy constructions of events, but we could. Since stating a construction of events that is faithful to what is known about human perception scientifically, or even to our common concepts of things, would put us in the business of formalizing a theory of perception, rather than a theory of the syntax-semantics interface for pictorial narratives, it is not advisable to go too far in this direction. But an analysis at the semantic level should make commitments about how normality enters into the semantics of the free perception construction.

The parts of the analysis that should be located somewhere are listed in (52). (52a) and (52b) were already located in the semantics of L_1. (52c) needs to be formalized, and it needs to be located somewhere in the semantic or pragmatic analysis.

(52) *Desiderata for the semantics of* L_1
   a. The agent picked out by discourse referent 1 is facing a scene like the second panel.
   b. That agent has just done a looking action e.
   c. e is an action that the agent could normally take while facing that scene.

The constraint in (52c) is related to circumstantial modality. This is the modal dimension that captures what is possible and what is normal in view of the state of the world. If we plant some hellebore seeds in the south yard, they might well germinate, and they might well not germinate. The soil is suitably moist and it's the right time of year for planting, but hellebores are finicky, the soil is overly acidic, and there are rodents that frequent the yard and tend to dig things up. If Boyer winks at an accomplice, Gable might not notice, even though Boyer's face is in Gable's field of view. If Boyer holds up the ace of spades, Gable will notice and see that it is the ace of spades. These semantics of modal constructions such as *might well* were captured by Kratzer (1981) in a semantic framework that, in examples of circumstantial modality, combines information about how things are in the base world (such as the facts about the soil and rodents) with information about what normally tends to happen given such-and-such facts. The first parameter is called a circumstantial modal base, while the second one is called an ordering source. In the hellebore example, the modal base consists of answers to questions such as *what is the soil pH?* and *how many animals of what species frequent the yard?*, while the ordering source might be propositions that encode a rule-of-thumb theory of horticulture.[53] In (52c), the modal base should consist of the answers to the question *what is the visual appearance of the scene the agent is facing?*, while

[53] Abusch (2012b) gives examples of ordering sources that capture rule-of-thumb physical theories. Where x is a tree, the following set captures that x normally does not fall, but if it falls, it can (in a normal course of events) fall in any direction other than West. This might support the truth of (i) and (iv), and the falsity of (ii) and (iii).

the ordering source should consists of propositions capturing a rule-of-thumb or scientific theory of what information the agent can normally pick up when he looks at such-and-such a scene.

Suppose we are given an event e, a possible world w, and an ordering source O. We want to use O to assess whether $w \diamond e$ is a normal evolution of world w. This is done by comparing $w \diamond e$ to competitors $w \diamond e'$ for optimality. In Kratzer's semantics, world u is strictly more optimal than world v relative to ordering source O if and only if the set of propositions from O that contain u is a proper superset of the set of propositions from O that contain v. If this condition is satisfied, u is more optimal than v in that moving from v to u adds propositions from O that are true. Definition (54) defines $u \diamond e$ to be a normal evolution of world u iff there is no competitor event e' such that $u \diamond e'$ is strictly more optimal than $w \diamond e$.[54]

(53) *Relative optimality of worlds*
World u is strictly more optimal than world v relative to O iff
$\{p | u \in p \wedge p \in O\} \supset \{p | v \in p \wedge p \in O\}$.

(54) *Normal evolution of a world*
Let u be a world, let e be an event such that $u \diamond e$ is defined, let E be a set of events, and let O be an ordering source. Then $u \diamond e$ is a normal evolution of u relative to E and O, written $\mathcal{N}(u, e, E, O)$, iff there is no $e' \in E$ such that $u \diamond e'$ is defined and is strictly more optimal than $u \diamond e$ relative to O.

Consider how to apply normality to a world of the form $w \diamond d_{\textbf{\textit{anc}}}$, where Gable has just looked at the scene with the ace, necktie, and open eyes, while attending to the ace, but not the necktie or Boyer's eyes. Let E be the set of events that are looking actions of Gable. This is used as the set of alternative events in definition (54). Accordingly, world $w \diamond d_{\textbf{\textit{anc}}}$ involves normal looking if there is no alternative $w \diamond e'$ where e' is a looking action of Gable that is strictly more optimal according to the ordering source than $w \diamond d_{\textbf{\textit{anc}}}$.

Should a normality entailment be included in the semantics of the free perception sequence (51)? Technically, it is possible to build a normality condition into the semantics of L_1, which is the covert seeing conjunct in extensional free perception. Including such a condition accounts for the fact that a reader of a narrative in which the sequence is included would assume that, in a described situation, the agent depicted in the first panels picks up that the card shown in the second panel is an ace. Similarly for a film which incorporates these frames. Further, it seems that authors

---

$$\left\{ \begin{array}{l} x \text{ does not fall,} \\ x \text{ either does not fall, or falls in a Northerly direction,} \\ x \text{ either does not fall, or falls in a Easterly direction,} \\ x \text{ either does not fall, or falls in a Southerly direction} \end{array} \right\}$$

(i) If the tree had fallen, it could have fallen in a Southerly direction.
(ii) If the tree had fallen, it could have fallen in a Westerly direction.
(iii) If the tree had fallen, it would have fallen in a Northerly direction.
(iv) If the tree had fallen, it would have fallen in a Northerly, Easterly, or Southerly direction.

[54]Definition (54) requires normal evolutions to be maximally optimal, in that there are no more optimal competitor events. See Kratzer (1981) for a more complex definition that deals with the possibility of infinite chains of increasingly more optimal worlds, without there being any maximally optimal worlds. This is potentially relevant in models with infinite sets of events.

of comics and creators of films intend for readers or viewers to draw such conclusions.

Normality assumptions are systematic in the interpretation of linguistic and pictorial narratives, though. In (55a), one automatically assumes that the key was used to open the lockbox by inserting it in the lock, rather than by prying at the hinge. In (55b), one assumes that the finger was broken by the impact of the softball, rather than some other cause. We think that in narratives, this kind of implication is well analyzed as an implicature to a stereotypical scenario (Atlas and Levinson, 1981; Horn, 1984). If the normality implication in extensional free perception sequences falls under this general phenomenon, one can claim that it is not necessary to write it into the semantics of L_1. On the other hand, formulating default theories and lexical entries is a common move in DRT (Kamp and Rossdeutscher, 1994; Asher and Lascarides, 2003). We leave this choice open, since our purpose in discussing normality is to motivate how it enters into the semantics of free-perception sequences that are understood intensionally.

(55)a. She opened the lockbox with a blue key.
b. Jack broke his right index finger while catching a softball.

This completes our definition of a revised model theory for perceptual events. The change from Section 3 is that there are normal events of Gable looking veridically at the scene with the ace, without picking up all of the information in the scene. We take it that the revised models reflect the ordinary understanding of viewers of films and readers of comics, as well as scientific understanding of human perception.

5. Weakening the embedding semantics

This section takes up the task of weakening the semantics of pictorial embedding, so that it is no longer subject to the problem of over-informative complements. The leading idea is that the epistemic consequences of events of hallucinating should be similar to the epistemic consequences of events of veridical looking, as modeled by event and world alternatives. For orientation, (56) lists events of veridical looking that Gable could participate in while facing the moon card scene, attending to at least the moon card, and not attending to everything. In the base world, these events have the same precondition of Gable facing a scene with a moon card (feature m), a necktie (feature n), and open eyes (feature o). Alternatives vary according to the attention features (annotated by bold face), with variation in the features that are not attended to. (57) provisionally describes a series of hallucinating events h^{m--}, h^{mn-}, and $h^{m,-,o}$ that have the same alternatives as $d_{\boldsymbol{mno}}$, $d_{\boldsymbol{mn}o}$, and $d_{\boldsymbol{m}n\boldsymbol{o}}$, respectively. h^{m--} is the event of Gable hallucinating a moon card, while not hallucinating anything specific about the neckwear or the eyes. The intended sense of "hallucinating a moon card" is that Gable gains the perceptual information that he is facing a moon card, though because he is hallucinating, he does so erroneously. What are the event alternatives for Gable? Ordinary veridical events of Gable viewing the moon card while attending to the card and not the neckwear or eyes are of the form $d_{\boldsymbol{m}--}$, where Gable looks at a scene with the moon card and attends to it. Filling in two possibilities for each of the open slots leads to the four alternatives shown on the first line in (56). All of the alternatives are taken to be events of veridical looking, and the alternatives for h^{m--} are the same as the alternatives for events of veridical looking of the form $d_{\boldsymbol{m}--}$, where

Gable looks at a scene with a moon card, and attends only to the card.[55]

(56)

event	precondition	alternatives	gloss
$d_{\boldsymbol{m}no}$	Gable faces moon, necktie, and open eyes	$\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}nc}$ $d_{\boldsymbol{m}so}, d_{\boldsymbol{m}sc}\}$	Gable looking veridically at the moon scene while attending only to the card
$d_{\boldsymbol{mn}o}$	Gable faces moon, necktie, and open eyes	$\{d_{\boldsymbol{mn}o}, d_{\boldsymbol{mn}c}\}$	Gable looking veridically at the moon scene while attending only to the card and neckwear
$d_{\boldsymbol{mno}}$	Gable faces moon, necktie, and open eyes	$\{d_{\boldsymbol{mno}}, d_{\boldsymbol{mso}}\}$	Gable looking veridically at the moon scene while attending only to the card and eyes

(57)

event	alternatives	gloss
$h^{m,-,-}$	$\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}nc}$ $d_{\boldsymbol{m}so}, d_{\boldsymbol{m}sc}\}$	Gable hallucinating a moon card
$h^{m,n,-}$	$\{d_{\boldsymbol{mn}o}, d_{\boldsymbol{mn}c}\}$	Gable hallucinating a moon card and necktie
$h^{m,-,o}$	$\{d_{\boldsymbol{mno}}, d_{\boldsymbol{mso}}\}$	Gable hallucinating a moon card and open eyes

Along the same lines, $h^{m,n,-}$ is glossed as the event of Gable hallucinating the moon card and a necktie on Boyer's neck, without hallucinating anything specific about the eyes. As shown on the second line of (57), this has the same alternatives as veridical looking events of the form $d_{\boldsymbol{mn}-}$, such as $d_{\boldsymbol{mn}o}$. This follows the design of hallucination events having epistemic consequences that parallel those of events of veridical looking. Event h^{m-o} is the event of Gable hallucinating a moon card and open eyes, while not hallucinating anything specific about the neckwear. Generalizing this, we form hallucinating events of the form h^{xyz}, where the feature values meet the conditions specified in (58). Here the attention features in veridical events are written as a bit vector, in place of boldfacing. For instance, d_{mno100} is the event $d_{\boldsymbol{m}no}$ of Gable looking at a configuration with moon card, a necktie, and open eyes, while attending only to the card.

(58) *Event alternatives for h^{xyz}*

Consider the event h^{xyz}, where in the first slot, x is either a (ace), m (moon), or $-$ (undetermined); in the second slot, y is either n (necktie), s (scarf), or $-$ (undetermined); and in the third slot, z is either o (open), c (closed), or $-$ (undetermined). The set of event alternatives to h^{xyz} is defined to be the set of events of the form $d_{x'y'z'tuv}$, where $x' \in \{a,m\}$, $y' \in \{n,s\}$, $z' \in \{o,c\}$, $t,u,v \in \{0,1\}$, $t = 0$ iff x is $-$, $u = 0$ iff y is $-$, $v = 0$ iff z is $-$, $x' = a$ if $x = a$, $x' = m$ if $x = m$, $y' = n$ if $y = n$, $y' = s$ if $y = s$, $z' = o$ if $z = o$, and $z' = c$ if $z = c$.

Section 3 introduced events such as the hallucinating event $h(g, q_m, q_a)$ of Gable facing a view just like the ace picture, while hallucinating the moon card scene. This raises the question whether events like h^{m--} should be split up into different events that have different preconditions about the scene that Gable is facing in the base world. We do not know whether this makes a difference. To be concrete, we will answer yes, and split h^{xyz} into events $h^{xyz}_{x''y''z''}$ that are parameterized in the

[55] The hyphens hava a different status in h^{m--} and $d_{\boldsymbol{m}--}$. In h^{m--} they are part of the notation for a certain event. In $d_{\boldsymbol{m}--}$ they are being used as schematic notation for four events, including $d_{\boldsymbol{m}no}$ and $d_{\boldsymbol{m}nc}$.

subscript by the scene that Gable is facing in the base world. The event alternatives are as specified in (58).

Definition (58) extends the model construction for the epistemic consequences of events with attention features to events of visual hallucination. When an agent views a scene veridically, the event in the base world records the features that the agent attends to. Event alternatives to that event are events of veridical looking that keep the features that are attended to constant, while other features can vary.[56] When an agent hallucinates, the visual features and feature values that are hallucinated are recorded in the event in the base world. Event alternatives are events of veridical looking, where the features that are recorded in the base event as hallucinated have the specified values in alternatives, and are attended to. For instance, h^{m--}_{ano} is the event of Gable hallucinating a scene with a moon card, while not hallucinating anything specific about the neckwear and eyes, and actually facing a scene with an ace, a necktie, and open eyes. The event alternatives are the set of veridical looking events $\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}nc}, d_{\boldsymbol{m}so}, d_{\boldsymbol{m}sc}\}$, where Gable looks veridically at a scene with a moon card, while the neckwear can be a scarf or a tie, and the eyes could be open or closed. The event h^{m--}_{ano} has weaker epistemic consequences than the event $h(g, q_m, q_a)$ from Section 3, which in current notation is h^{mno}_{ano}.

Now we are ready to reformulate the semantics of the intensional free perception narrative repeated in (59). The earlier model construction and syntax-semantics interface had the consequence that all alternatives to a world $w \diamond e$ that satisfies the formula had Gable facing a scene just like the moon picture. This was problematic, because it has Gable gaining too much information, and moreover has him gaining more information when he hallucinates than he normally would when looking veridically.

(59)  $a \; S_1$ 

Defining the semantics of LF (59) is posed as the problem of defining whether the formula is true, false, or undefined in an arbitrary world of the form $w \diamond e$, relative to an arbitrary witness sequence $\mathscr{O}$. (60) lists some choices for the event e in $w \diamond e$, and the truth values which will be obtained below. The LF is undefined in $w \diamond c$, because c is an event of Gable coughing, rather than a looking event, and so the presupposition of S_1 is not satisfied. Event h^{m--}_{ano} is the base event of Gable hallucinating a moon card, and the formula will come out true in a world of form $w \diamond h^{m--}_{ano}$. Event h^{-s-}_{ano} is a base event of Gable hallucinating a scarf, in a base situation where he faces an ace, necktie, and open eyes. In a way that will be explained below, this event does not have epistemic consequences that are compatible with the embedded picture, and formula (59) comes out false in a world ending with this event. Event $d_{a\boldsymbol{n}o}$ is the event of Gable veridically attending to the necktie in a situation where he faces an ace, necktie, and open eyes. For a subtle reason, (59) will come out false in a world ending with this event.

[56] While the parts of event symbols such as $d_{\boldsymbol{m}no}$ have certain intended interpretations, in the model structure this interpretation is mediated by the preconditions and the event alternatives. Symbols such as $d_{\boldsymbol{m}no}$ might as well be atomic.

(60)

event	truth value	description
c	undefined	Gable coughing
h^{m--}_{ano}	true	Gable hallucinating a moon card
h^{-s-}_{ano}	false	Gable hallucinating a scarf
$d_{\boldsymbol{ano}}$	false	Gable veridically attending to the neckwear in a situation where he faces an ace, necktie, and open eyes

The semantics of $S_1(q)$ should be spelled out with reference to the semantic value of the embedded picture q, and to the world alternatives to the base world $w \diamond e$. The world alternatives are determined by event alternatives to e. For base events from (60), (61) lists the event alternatives for Gable.

(61)

event	event alternatives
c	$\{\mathrm{c}\}$
h^{m--}_{ano}	$\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}nc}, d_{\boldsymbol{m}so}, d_{\boldsymbol{m}sc}\}$
h^{-s-}_{ano}	$\{d_{a\boldsymbol{s}o}, d_{a\boldsymbol{s}c}, d_{m\boldsymbol{s}o}, d_{m\boldsymbol{s}c}\}$
$d_{\boldsymbol{ano}}$	$\{d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}, d_{a\boldsymbol{n}o}, d_{a\boldsymbol{n}c}\}$

Looking at the second line, updating the world alternatives to the base world $w \diamond h^{m--}_{ano}$, where Gable has just hallucinated the moon card, results in world alternatives of the form $w_1 \diamond d_{\boldsymbol{m}no}$, $w_2 \diamond d_{\boldsymbol{m}nc}$, $w_3 \diamond d_{\boldsymbol{m}so}$, and $w_4 \diamond d_{\boldsymbol{m}sc}$, where $w_1, ..., w_4$ are world alternatives to w. A Hintikka semantics for complementation would check that each of these worlds satisfies the embedded picture, in the sense that the world looks like the embedded picture from Gable's geometric perspective. This will not work, because in a world of the form $w_4 \diamond d_{\boldsymbol{m}sc}$, Gable is facing a scene where Boyer is holding up a moon card while wearing a scarf and having closed eyes in w_4.[57] This information is imposed by the preconditions of event $d_{\boldsymbol{m}sc}$, and the information is inconsistent with the content of moon card picture, because in that picture, Gable is wearing a necktie rather than a scarf, and has open rather than closed eyes. This situation comes from event h^{m--}_{ano} having more alternatives than event $h(g, q_a, q_m)$ from Section 3, so that universal quantification is harder to satisfy.

The source of the problem is simply that in the event model, the epistemic consequences of most hallucinating events, which are adapted from the epistemic consequences of events of veridical looking, are not strong enough to entail the content of embedded pictures.[58] A tentative solution is to adjust quantificational force: there is *some* world alternative to $w \diamond e$ that supports the content of the embedded picture. Against the background of Hintikka semantics for embedding, this move is surprising. And as stated, it is unacceptably weak. Take the event $d_{\boldsymbol{ano}}$ of Gable looking veridically at a view with an ace, a necktie, and open eyes and attending to the neckwear. As seen in the last line of (61), this event has alternatives $\{d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}, d_{a\boldsymbol{n}o}, d_{a\boldsymbol{n}c}\}$, and of these $d_{m\boldsymbol{n}o}$ satisfies the moon card picture.[59] It follows that on the weak analysis with existential quantification, formula (59) is true in the world of the form $w \diamond d_{\boldsymbol{ano}}$ where Gable faces an ace, necktie, and open eyes and looks

[57] This phrasing comes from the fact that correspondence to the scene is stated in the pre-conditions of the event, so that it is w_4 that has to agree with the embedded picture. If event $d_{\boldsymbol{m}sc}$ does not change the scene, the distinction does not matter.

[58] Or especially, of film shots. The argument being made here is particularly compelling for film, because their geometric, temporal content (and for sound film, acoustic content) is so strong.

[59] This is a shorthand which is expanded as any defined world of the form $u \diamond d_{m\boldsymbol{n}o}$ being such that u looks like the picture from the geometric visual perspective of the agent.

veridically while attending to the neckwear. This seems unacceptable. If we stipulate that Gable is not hallucinating, which he isn't in $w \diamond d_{\boldsymbol{a}no}$, then (59) should be false in $w \diamond d_{\boldsymbol{a}\boldsymbol{n}o}$, but it comes out true in the weak formulation. This indicates that the requirement of some world alternative to $w \diamond e$ supporting the content of the embedded picture is too weak.

We trace the defect to a failure to refer to normality in the semantics for intensional embedding. In a scenario where Gable and Boyer are playing cards, when Boyer holds up a card, Gable can be expected to attend to it. This should make the world $d_{\boldsymbol{a}\boldsymbol{n}o}$, where Gable attends only to the neckwear, irrelevant as a witness to the truth of the narrative. This is remedied by including a normality condition: the narrative is true in $w \diamond e$ if and only if there is some world alternative $u' \diamond e'$ to $w \diamond e$ for Gable that satisfies the embedded picture (i.e. where u' looks like the moon card picture from Gable's geometric perspective), and where e' is a looking action that Gable could *normally* take while facing the visual scene that he is facing in u'. This definition is tested and illustrated by considering base world events, their event alternatives, and witness worlds for the truth of the moon card narrative. In (62), h^{m--}_{xyz} is an event of Gable hallucinating a moon card (and no other features) in a base world where he is facing a scene with card x, neckwear y, and eye configuration z. The base world is given in this schematic form because the base world features are not relevant. This event has the event alternatives $\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}\boldsymbol{n}c}, d_{\boldsymbol{m}\boldsymbol{s}o}, d_{\boldsymbol{m}\boldsymbol{s}c}\}$, and of these $d_{\boldsymbol{m}no}$ is a perceptual action that Gable can take while facing the moon card scene. It is normal on the assumption that when playing cards and a card is held up, Gable can attend just to the card and not the neckwear or the eyes in a normal course of events. Therefore a world of the form $w_1 \diamond d_{\boldsymbol{m}no}$ is a witness for the existential modal requirement.

(62) Base world events, event alternatives, and modal world witnesses for moon card narrative, assuming an ordering source where it is normal for Gable to attend to the card whenever a card is held up.

event	event alternatives	witness world for (59)
h^{m--}_{xyz}	$\{d_{\boldsymbol{m}no}, d_{\boldsymbol{m}\boldsymbol{n}c}, d_{\boldsymbol{m}\boldsymbol{s}o}, d_{\boldsymbol{m}\boldsymbol{s}c}\}$	$w_1 \diamond d_{\boldsymbol{m}no}$
h^{mn-}_{xyz}	$\{d_{\boldsymbol{m}\boldsymbol{n}o}, d_{\boldsymbol{m}\boldsymbol{n}c}\}$	$w_2 \diamond d_{\boldsymbol{m}\boldsymbol{n}o}$
h^{m-o}_{xyz}	$\{d_{\boldsymbol{m}\boldsymbol{n}\boldsymbol{o}}, d_{\boldsymbol{m}\boldsymbol{s}\boldsymbol{o}}, \}$	$w_3 \diamond d_{\boldsymbol{m}\boldsymbol{n}\boldsymbol{o}}$
h^{-s-}_{xyz}	$\{d_{\boldsymbol{a}\boldsymbol{s}o}, d_{\boldsymbol{a}\boldsymbol{s}c}, d_{m\boldsymbol{s}o}, d_{m\boldsymbol{s}c}\}$	*none*
h^{-n-}_{xyz}	$\{d_{\boldsymbol{a}\boldsymbol{n}o}, d_{\boldsymbol{a}\boldsymbol{n}c}, d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}\}$	*none*
$d_{\boldsymbol{a}\boldsymbol{n}o}$	$\{d_{\boldsymbol{a}\boldsymbol{n}o}, d_{\boldsymbol{a}\boldsymbol{n}c}, d_{m\boldsymbol{n}o}, d_{m\boldsymbol{n}c}\}$	*none*

On the second line of (62), h^{mn-}_{xyz} is an event of hallucinating a moon card and a necktie. The increase in attention results in two alternatives $d_{\boldsymbol{m}\boldsymbol{n}o}$ and $d_{\boldsymbol{m}\boldsymbol{n}c}$ rather than four. These still include an option (the first) which is consistent with the embedded picture, and a world of the form $w_2 \diamond d_{\boldsymbol{m}\boldsymbol{n}o}$ is a witness for the truth of the narrative, assuming attending to both the card and the neckwear is normal. The third line is similar. On the fourth line, h^{-s-}_{xyz} is an event of hallucinating a scarf. For this the event alternatives all involve seeing a scarf, and since the moon card picture shows a tie, no alternative world satisfies the picture. Therefore the narrative is false in a world with a base event h^{-s-}_{xyz}. On the fifth line, h^{-n-}_{xyz} is an event of hallucinating a necktie. This again has four event alternatives, which this time includes an event $d_{m\boldsymbol{n}o}$ that is consistent with the embedded picture. But it is ruled out by normality, on the assumption that in a card game, Gable would normally attend to a card when one is held up. So there is no witness, and the narrative is false in a world where h^{-n-}_{xyz} has just happened. On the sixth line, $d_{\boldsymbol{a}\boldsymbol{n}o}$ is an event of veridically looking at an

ace, necktie, and open eyes, while attending only to the neckwear. Here too there is an alternative that satisfies the embedded pictures. It is ruled out by normality, just as in line 5. Lines 5 and 6 illustrate that adding normality strengthens the semantics, in the sense that the narrative becomes false in some worlds where the existential condition without normality is satisfied.

There is another place in pictorial materials where a semantics of existential modal quantification with a normality condition is motivated. Consider a grade school history textbook which presents as factual the scenario of Norse explorers led by Leif Erikson having visited the North American continent. The text is accompanied by the illustration in (63) (Wergeland, 1884). A commentator could object that while the evidence for Norse settlement in North America is good, coming from Icelandic historical tales and archeological findings in the L'Anse aux Meadows area of Newfoundland, the historical record is not at all strong enough to support the proposition that the Norse landing in North America looked exactly like (63). So the textbook is inaccurate and is known by the authors to be inaccurate. A response is that the illustration is intended as an "artist's impression", something that is maintained to be consistent with the historical record, but which need not be entailed by the historical record in order for the textbook to be accurate. Consider a variant of the picture where the large figure in the center right, instead of carrying a spear, is shown carrying a Roman staff with the inscription SPQR. It is in fact consistent with the historical record that a Norseman carried a Roman staff with the inscription SPQR to Newfoundland. But an illustration like this would not be included in a grade school history textbook. The reason is that artist's impressions in history textbooks should present visual information that is not only consistent with the historical record, but normal given the historical record and practices of historical interpretation. Technically, the force of the illustration in the textbook is that there is a world w' that is consistent with the historical record, where w' looks like the picture from a vantage point in the area of north-eastern North America in the late middle ages, and where w' is a normal world relative modal parameters that capture the historical record and standards of historical interpretation.[60] In terminology suggested by this analogy, the analysis of pictorial free perception developed here, instead of requiring that the attitudinal content of the hallucinating event in the base world entails the content of the embedded picture, treats the embedded picture as an "artist's impression" of what a world consistent with the attitudinal content of the event in the base world could have looked like, under a constraint of normality.

[60]We guess that the historical record should be modeled in the modal base, and standards of historical interpretation in the ordering source.

(63)

The discussion of normality in the Gable-Boyer scenario referred to the assumption that in a card game, if a card is held up, Gable would attend to it in any normal course of events. This is dependent on the fact that Gable and Boyer are playing cards in the base world. This kind of dependence is built into the semantics for modality as formulated by Kratzer, where the ordering source parameter is a function from possible worlds to sets of propositions (Kratzer, 1978, 1981). So in this framework, it is possible to express that different looking events are normal in different situations.

Formulating the analysis with normality is straightforward, because everything can be packed into the semantics of the hidden seeing syntax S_1. As before, we state the semantics by defining a semantic relation, and then define satisfaction conditions for the syntactic embedding predicate S_1 in terms of it. $\widetilde{S}(w,x,Q)$ presupposes that w finishes with a looking action e by agent x. The intensional entailment is that some alternative to u for x is of the form $u' \diamond e'$, where u' and the geometric viewpoint of x in u' satisfy the pictorial content Q, and where $u' \diamond e'$ is a normal development of u' according to the ordering source. The latter appears as $O(w)$, because the ordering source parameter is a function from worlds to sets of propositions. This makes the semantics sensitive to the standards of normality that prevail in the base world. The definition is (64).

(64) *Semantics of pictorial embedding relation $\widetilde{S}(w,x,Q)$, final event version.*

Let w be a world, x an individual, and Q a relation between worlds and viewpoints.

Presupposition: w is of the form $w' \diamond e$, where e is a looking action with agent x.

Assertion: $\exists u[R(w,x,u) \wedge \exists u' \exists e'[u = u' \diamond e' \wedge Q(u', V(u',x)) \wedge \mathscr{N}(u,e',E,O(w))]$, where E is the set of events of x looking, O is an ordering source function modeling normal looking, and R is the epistemic alternative relation.

It is significant to note that while the definition of the embedding semantics uses existential force, when a base world w is incremented with event h^{m--}_{ano} (the event of Gable hallucinating a moon card), all of the world alternatives to Gable in $w \diamond h^{m--}_{ano}$ have Gable facing a scene with a moon card. Each alternative is of the form $u' \diamond e'$, where e' is an alternative to h^{m--}_{ano}. And as seen in (57), each such alternative has Gable as the agent of an event of looking at a scene with a

moon card. This is information is imposed on u' via preconditions. So after h^{m--}_{ano} happens, Gable will believe that Boyer is holding up or has held up a moon card, in the sense that this is true in all epistemic alternatives. Although the syntax-semantics interface uses existential quantification, Gable is described as picking up the information that there is a moon card. The route to this is indirect, via the event alternative construction, preconditions of veridical looking events, and the normality condition. At the same time, not all of Gable's alternatives are required to look exactly like the embedded picture from Gable's visual perspective, in order for LF (59) to be satisfied. This is the subtle way in which the existential semantics, as it combines with normality and the event-alternative epistemic semantics, weakens the embedding semantics without losing the entailment that Gable is described as picking up the information that Boyer is holding up a moon card.

6. Discussion

This section takes stock of the results of the investigation, and mentions some extensions and alternative analyses. A general result is that it is possible to concretely formulate a possible worlds semantics for free perception sequences in pictorial narratives, in a framework where pictures have the semantics of relations between worlds and viewpoints. Free perception sequences are implicitly anaphoric, with reference in the interpretation of the second picture to an agent depicted in the first one. This is formulated in a straightforward way in a dynamic framework, using implicit syntax that introduces discourse referents for depicted individuals, and implicit operators that pick up referents for discourse referents.

Section 2 introduced an LF using L_1 for extensional readings of free perception, and an LF using $S_1(q)$ for intensional readings. Is there reason to believe that extensional and intensional readings are grammatically distinct? It is inevitable that narratives that are understood to describe hallucination, imagination, or memory have embedding operators, because in a non-embedded logical form, entailments of top-level conjuncts survive as entailments of the narrative as a whole. But having formulated a semantics for the embedding construction, we should ask whether a grammar that also includes a distinct extensional operator is motivated. Section 2 took the stance that extensional readings are continuous with general extensional narratives, with extensional free perception readings resulting from adding some information to the general semantics. It brought up other narrative patterns (over-the shoulder shots, sightlines and eye reflections) where a grammar of embedding is implausible, and which are parallel in their entailments to extensional readings of free perception sequences. This was claimed to favor a grammar that included L_1. But if it is possible to assume the syntax $S_1(q)$ for both extensional and intensional "readings" of free perception sequences, this would be attractive on grounds of parsimony.[61]

The question then is whether it is possible to start with the syntax $S_1(q)$, and then by some motivated process add information to arrive at the equivalent of $q\,L_1$. $S_1(q)$ already presupposes that a verifying world is of the form $u \diamond e$, where e is a looking action by the agent picked out with index 1. What if this basic meaning was strengthened or "enriched" to e being a *veridical* looking action by the agent picked out by index 1? This is natural, because veridical looking is

[61] An L&P reviewer brought up a different, intriguing idea. Given that as discussed in Section 4, there are a variety of intensional operators in film which in their basic form are not point of view constructions, such as imagination and memory, it could be attractive to split S_1 into an intensional component and a point of view component. Then L_1 or a version of it would be included as the point of view operator, and intensionality would be added by a distinct operator that is also used elsewhere.

a default case. In the event model as defined in Sections 4 and 5, the veridical enrichment would exclude e being a hallucinating event such as h^{m--}_{ano}, and would restrict e to being an ordinary looking event such as $d_{\boldsymbol{a}no}$ or $d_{a\boldsymbol{n}o}$. Is $S_1(q)$ plus the veridical enrichment equivalent to $q\,L_1$? In particular, does $S_1(q)$ with the veridical enrichment entail that the base world looks like q from the visual-geometric perspective of the agent? Given our semantics and model theory, it does not. Consider a base world where Gable is facing a scene with an ace, a necktie and *closed* eyes, where the event $d_{\boldsymbol{a}nc}$ of looking at such a scene while attending to the card has just happened. The base world event has alternatives $\{d_{\boldsymbol{a}no}, d_{\boldsymbol{a}nc}, d_{\boldsymbol{a}so}, d_{\boldsymbol{a}sc}\}$. Of these $d_{\boldsymbol{a}no}$ is consistent with the ace picture, which shows Boyer holding up an ace while wearing a necktie and with open eyes. Moreover $d_{\boldsymbol{a}no}$ is normal, because it has Gable attending to the card. All of this means that the narrative repeated in (65), with an enrichment to veridical looking, can be true in a world where Gable is facing closed eyes. This shows that $S_1(q)$ with the veridical enrichment does *not* entail that the base world looks like q from the agent's perspective. So it is not possible to eliminate $q\,L_1$ in favor of $S_1(q)$ plus the enrichment that the agent is looking veridically. Of course, this result is tied closely to our semantics and model theory. And perhaps there is another pragmatic route to the veridical reading. We do take it that it is desirable to provide for a reading (possibly derived by implicature or enrichment) which entails that the base world looks like the second picture in a free perception sequence, since this is the basic and most common case in pictorial narratives.

(65)  $a \quad S_1$ 

Sections 2 and 3 presented an initial model theory and semantics where embedding LFs describe agents as having alternatives that look exactly like the embedded picture. This was criticized in Section 4 as attributing too much information to the agent. This point is particularly strong in examples from film, because the geometric and temporal information in a film shot is so strong. Would it be possible after all to stick with the original analysis? There are a couple of points which might make this stance more plausible. It could be that agents initially have very rich visual and acoustic information, as encoded in their attitudinal alternatives, but quickly forget much of the information. This would shield the analysis from Sections 2 and 3 from the consequence that agents are described as picking up such rich geometric and acoustic information. The agent would have the information only temporarily, and not gain the information permanently. Second, agents could have rich geometric information, without being able to draw consequences from it. In the scenario of the locusts, both in the veridical and non-veridical versions, perhaps the agent Keisha is presented as having an experience that is geometrically as rich as the picture with the locusts, but is not able to draw some consequences. In particular, she is not able to draw the consequence that there are two hundred and eleven locusts. The notion that human agents temporarily get all of the information in their visual scenes, and that when they hallucinate, imagine, or remember they temporarily access the same volume of information is certainly scientifically false. But perhaps it is a piece of folk psychology, which is exploited in pictorial narratives. We wonder though what the point would be in constructing models where agents temporarily get massive amounts of information when they look, hallucinate, remember, or imagine, (as captured by attitudinal

alternatives), without their having access to the information, without the information influencing their behavior, and only to immediately forget most of it. At the limit, one could claim that agents access a unique world when they look, hallucinate, remember, or imagine, and that images or shots in attitudinal depictions in pictorial narratives are images or shots of that unique world.[62] But they don't have access to properties of that world, and immediately forget which world it is. Even if this kind of move is technically feasible, it is arguably unattractive, because the strong information that it attributes to agents is not motivated by considerations of modeling the attitudinal state of agents. Rather, it imports into the model of the attitudinal state of agents the specificity of the picture or film shot. Consider a free perception sequence in a virtual reality narrative, where the point of view "shot" includes detailed three-dimensional information. This is extra information, relative to an ordinary film shot. The strategy of importing information into the possible worlds model of the attitudinal state of agents would dictate that an agent in the base world has the detailed three-dimensional geometric information, in the sense that in attitudinal alternatives, the agent is in a world with the same local three-dimensional geometry. This change would be motivated not by some discovery about agents, but by a new kind of narrative having been invented. We prefer to work with models that are motivated by modeling the state of agents, and deal in the syntax-semantics interface with the mismatch between the information of agents and the information in embedded pictorial narratives. But the alternative analysis that attributes strong information to agents strikes us as feasible enough that it merits further consideration.

The analysis of over-informative complements in Section 5 was phrased as a definition of an embedding relation $\widetilde{S}$, which is used in formulas $\widetilde{S}(w,x,Q)$, where Q is a pictorial content. This formalizes the semantics free-perception sequences in pictorial narratives that describe non-veridical seeing. $\widetilde{S}$ is weaker in the information that it attributes to agents than Hintikka semantics for embedding. Sections 4 and 5 mentioned other pictorial materials that motivate a weakened semantics, namely film shots depicting imagination or memory, and historical artist's impressions. These examples figured in the argument, in that they are further examples of over-informative complements. But their semantics is not formalized, or in the case of artist's impressions, only sketched. We conjecture that weakened quantificational force and a normality condition should also be included in the semantics of these constructions, in order to deal with over-informativity.[63]

There is another place in recent literature where an apparently over-informative pictorial contents come up. Esipova (2021) discusses the phenomenon of pictures in instructional and prohibition signs making a weakened compositional contribution. The prohibition sign at the top in (66) is interpreted in a way that is equivalent to the text at the bottom, not as something along the lines of "coyotes may not be fed from above" (Keith, 2007). Esiposa theorizes that the pictorial content is divided into at issue and not-at-issue components, and that only the at issue part contributes to the modal prohibition. This is potentially a different way of weakening the semantics of intensional free perception sequences. In the moon card sequence, one could postulate that only the information about the card that is held up is at issue, so only this information makes a contribution to the attitudinal semantics. Or potentially, at-issueness is a different way of glossing the information in

[62] Since replacing a set of world alternatives with a unique "belief world" does not encode the information in a world-alternative relation, world-alternative relations characterizing the epistemic states of agent would be needed too. This makes the point that the extra information is motivated by the narrative, not modeling the epistemic states of agents. A way of conceptualizing this strategy is that it imports into the base world the witness world u in (64) that projects to the embedded picture.

[63] See Section 4 of Abusch and Rooth (2022).

the ordering source in our formulation.[64]

(66) 

The event models in this paper in their technical form are sandbox models, which are motivated by the desire to explore ideas in detail. They are idealized in several ways. In more realistic event models as they are used in linguistic semantics, multiple events can happen at once, events can partially overlap, and events have a part-whole structure. Independently, the modeling of hallucinating events in Section 4 was confined to scenarios of "ordinary hallucination", where the agent gets certain non-actual information and accepts it. But it is also possible to have a visual experience that one knows to be non-real, or suspects might not be real.[65] One of the authors has had the visual experience of the trees in his back yard moving backwards and forwards against the background of a sunset. The experience is pleasant and interesting rather than scary, and it does not induce him to get out of the way. This kind of example calls for an elaboration of epistemic event models to allow for known and suspected hallucination, while maintaining a connection between ordinary hallucination, known or suspected hallucination, and events of veridical looking. If the formulations in this paper are attractive, one would certainly like to know whether the various idealizations can be lifted, and whether elaborations in the epistemic semantics can be accommodated. We do not know whether this might undermine our results.

References

Abrusán, M. (2021). The spectrum of perspective shift: protagonist projection versus free indirect discourse. *Linguistics and Philosophy 44*(4), 839–873.

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

Abusch, D. (2012b). Circumstantial and temporal dependence in counterfactual modals. *Natural Language Semantics 20*(3), 273–297.

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. (2022). A possible-worlds construal of unreliability in film. *Epistemology & Philosophy of Science*, 38–42.

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

[64]Thanks to an L&P reviewer from bringing up this connection.

[65]This was pointed out by an L&P reviewer.

Abusch, D. and M. Rooth (2021). Modalized normality in pictorial narratives. In *Proceedings of Sinn und Bedeutung*, Volume 25, pp. 1–18.

Abusch, D. and M. Rooth (2022). Temporal and intensional pictorial conflation. In *Proceedings of Sinn und Bedeutung 26*.

Alberti, L. (1435). *Della Pittura*.

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

Atlas, J. D. and S. C. Levinson (1981). It-clefts, informativeness and logical form. In *Radical Pragmatics*, pp. 1–62. Academic Press.

Baltag, A., L. S. Moss, and S. Solecki (1998). The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, pp. 43–56.

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge.

Barry, C. (1975). *Poldark*. BBC.

Barwise, J. (1981). Scenes and other situations. *The Journal of Philosophy 78*(7), 369–397.

Bell, C. (2014). *El Deafo*. Abrams.

Bilal, E. (1992). *Froid Équateur*. Les Humanoïdes Associés.

Bimpikou, S. (2018). Perspective blending in graphic media. *ESSLLI 2018 Student Session*, 245–258.

Bittner, M. (2001). Surface composition as bridging. *Journal of Semantics 18*(2), 127–177.

Bordwell, D. (2012). *Poetics of Cinema*. Routledge.

Branigan, E. (1975). Formal permutations of the point-of-view shot. *Screen 16*(3), 54–64.

Branigan, E. (1984). *Point of View in the Cinema*. De Gruyter Mouton.

Brinton, L. (1980). 'Represented perception': A study in narrative style. *Poetics 9*(4), 363–381.

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.

Chabris, C. F. and D. J. Simons (2010). *The Invisible Gorilla*. Harmony.

Cuarón, A. (2013). *Gravity*. Warner Brothers Pictures.

Cumming, S., G. Greenberg, E. Kaiser, and R. Kelly (2021). Showing seeing in film. *Ergo 7*, 730–756.

Dekker, P. (1994). Predicate logic with anaphora. In *Proceedings from Semantics and Linguistic Theory*, Volume 4, pp. 79–95.

Dekker, P. J. (2012). *Dynamic Semantics*, Volume 91. Springer Science & Business Media.

Delgado, R. (1997). *Age of Reptiles: The Hunt*. Dark Horse Comics.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse: How texts allow us to mind-read and eavesdrop*. Brill.

Esipova, M. (2021). On not-at-issueness in pictures. *Glossa: a Journal of General Linguistics 6*(1), 83.

Felser, C. (1999). *Verbal Complement Clauses: A Minimalist Study of Direct Perception Constructions*, Volume 25. John Benjamins Publishing.

Gilliam, T. (1998). *Fear and Loathing in Las Vegas*. Universal Pictures.

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

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. *Perception 5*, 89–110.

Gross, L. (2006). Living out loud: Lost in Over-the-Rhine. *CityBeat*.

Hart, B. (2014). Ghosts. *Granta*.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. thesis, University of Massachusetts, Amherst.

Heim, I. (1987). Where does the definiteness restriction apply? evidence from the definiteness of variables. *The Representation of (In)definiteness 14*, 21–42.

Higginbotham, J. (1983). The logic of perceptual reports: An extensional alternative to situation semantics. *The Journal of Philosophy 80*(2), 100–127.

Hinterwimmer, S. (2017). Two kinds of perspective taking in narrative texts. In *Semantics and Linguistic Theory*, Volume 27, pp. 282–301.

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

Horn, L. (1984). Towards a new taxanomy for pragmatic inferences: Q-based vs. R-based implicatures. In *Georgetown Round Table on Languages and Linguistics*.

Howard, R. (2001). *A Beautiful Mind*. Universal Pictures and Dreamworks Pictures.

Kamp, H. (1979). Events, instants and temporal reference. In *Semantics from Different Points of View*, pp. 376–418. Springer.

Kamp, H. (1981). A theory of truth and semantic representation. In *Prodeedings of the 3rd Amsterdam Colloquium*, pp. 277–322.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Springer.

Kamp, H. and A. Rossdeutscher (1994). DRS-construction and lexically driven inference. *Theoretical Linguistics*.

Keith, J. (2007). Do not feed the coyotes. `https://www.flickr.com/photos/adactio/2156791720/`. CC BY 2.0.

Kozen, D. (2001). Automata on guarded strings and applications. Technical report, Cornell University.

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

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

Krifka, M. (1989). *Nominalreferenz und Zeitkonstitution: Zur Semantik von Massentermen, Pluraltermen und Aspektklassen*. W. Fink.

Lawrence, D. H. (1913). *Sons and Lovers*. Duckworth.

Lewis, D. (1975). Adverbs of quantification. In *Formal Semantics of Natural Language*, pp. 178–188. Cambridge University Press.

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

Lewis, D. (1986). *On the plurality of worlds*. Blackwell Oxford.

Lia, S. (2007). *Fluffy*. Jonathan Cape.

Liefke, K. (2022). The filmic representation of 'relived' experiences. *Epistemology & Philosophy of Science*, 56–65.

Liu, L., W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen (2020). Deep learning for generic object detection: A survey. *International Journal of Computer Vision 128*(2), 261–318.

Maier, E. (2019). Picturing words: the semantics of speech balloons. In *Proceedings of 22nd Amsterdam Colloquium*.

Maier, E. (2022). Unreliability and point of view in filmic narration. *Epistemology & Philosophy of Science*, 23–37.

Maier, E. and S. Bimpikou (2019). Shifting perspectives in pictorial narratives. In *Proceedings of Sinn und Bedeutung 23*, pp. 1–15.

McCarthy, J. (1963). Situations, actions, and causal laws. Technical report, Stanford CS.
Nelles, W. (2002). *Stories Within Stores: Narrative Levels and Embedded Narrative*. Columbus: Ohio State University Press.
Nelles, W. (2010). Embedding. In D. Herman, M. Jahn, and M.-L. Ryan (Eds.), *Routledge Encyclopedia of Narrative Theory*. Routledge.
Reed, C. (1949). *The Third Man*. London Films.
Reiter, R. (2001). *Knowledge in Action: Logical foundations for specifying and implementing dynamical systems*. MIT press.
Rooth, M. and D. Abusch (2019). Indexing across media. In *Proceedings of 22nd Amsterdam Colloquium*. ILLC, University of Amsterdam.
Schlenker, P. (2004). Context of thought and context of utterance: A note on free indirect discourse and the historical present. *Mind & Language 19*(3), 279–304.
Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy 31*(3), 353–395.
Simons, D. J. and C. F. Chabris (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception 28*(9), 1059–1074.
Simons, D. J. and D. T. Levin (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review 5*(4), 644–649.
Stiller, M. (1927). *The Woman on Trial*. Paramount Pictures.
Tan, S. (2006). *The arrival*. Hodder Children's Books.
Tezuka, O. (2015). *Ode to Kirihito*. Kodansha USA.
Thomason, R. H. (1984). Combinations of tense and modality. In *Handbook of Philosophical Logic*, pp. 135–165. Springer.
Thompson, H. S. (1972). *Fear and Loathing in Las Vegas*. Random House.
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.
Van Ditmarsch, H., W. van Der Hoek, and B. Kooi (2007). *Dynamic Epistemic Logic*. Springer.
Wergeland, O. (1884). *Leif Ericson Off the Coast of Vineland*. Selmar Hess.
Zeman, A., M. Dewar, and S. Della Sala (2016). Reflections on aphantasia. *Cortex 74*, 336–337.
Zeman, A., F. Milton, S. Della Sala, M. Dewar, T. Frayling, J. Gaddum, A. Hattersley, B. Heuerman-Williamson, K. Jones, M. MacKisack, et al. (2020). Phantasia–the psychological significance of lifelong visual imagery vividness extremes. *Cortex 130*, 426–440.
Zimmer, C. (2021). Many people have a vivid 'mind's eye,' while others have none at all. *New York Times*.
Zimmermann, T. E. (2008). Kontextabhängigkeit. In *Semantik/Semantics*, pp. 156–229. De Gruyter Mouton.