Modalized normality in pictorial narratives[1]

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

Abstract. An analysis of pictorial free perception using universal quantification over worlds that are consistent with the embedded picture is incorrect, because it entails that depicted characters pick up all of the information in a scene they look at. Human agents typically do not. This is corrected by replacing universal quantification over epistemic alternatives with existential quantification, and by introducing a normality condition.

Keywords: Free indirect disourse, free perception, perceptual blindness, intensionality, possible worlds semantics, event semantics, comics, film, compositional semantics.

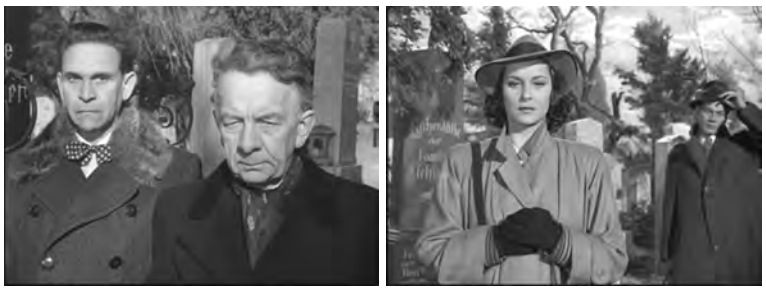
1. Introduction

Free perception sequences are ones where a setup sentence in a linguistic narrative, a setup panel in a comic, or a setup shot in a film shows a character looking, and the subsequent sentence, panel, or shot is understood as describing or showing what the character sees. (1) is a linguistic example, where the first clause describes the first-person character looking, and the second clause is understood to convey that they saw an angry kid approaching them.

(1) I looked up, and that angry kid was walking toward me.

(2) shows two frames from the film *The Third Man*, corresponding to two shots. The end of the first shot shows a man with a fur collar looking off camera to his left, and the second shot is understood as showing what he sees.

(2)



(3) is an example from Simone Lia's *Fluffy*. Michael has lost his rabbit on a train. Searching, he looks into a cabin and he sees a girl eating a rabbit sandwich. It is later clarified that he was hallucinating–the girl was eating a kipferl, a kind of Austrian pastry. This example shows that free perception sequences can be intensional, in that they describe an agent's perceptual and belief state, rather than conveying information about basic facts in a described situation.

[1]We thank participants in the Fall 2020 seminar on supersemantics at Cornell for their comments, and we thank participants in Sinn und Bedeutung 25 for their comments and attention. Images that are quoted from comics and film are used for educational and critical purposes, and are property of their respective owners.

 In: P. G. Grosz, L. Martí, H. Pearson, Y. Sudo, and S. Zobel (eds.) *Proceedings of Sinn und Bedeutung 25*, pp. 1–18. University College London and Queen Mary University of London.

(3) 

Abusch and Rooth (2017) analyzed free perception sequences in a possible worlds framwork. The analysis uses an analysis of epistemic states in Kripke frames, where the epistemic state of an agent is captured using a world-alternative relation. Intensional free-perception sequences are analyzed as involving syntactic embedding. The analysis can be characterized as giving a semantic analyis of pictorial intensional free perception that is modeled on clausal embedding in language, as in (4). Effectively, the fluffy panel has the syntax (5), and the embedding of the panel is analyzed semantically as intensional embedding, using universal quantification over alternatives.

(4) a. He saw that an angry kid was walking toward him.
b. He believed that an angry kid was walking toward him.

(5) He saw  .

This kind of analysis using covert embedding is discussed for natural language free indirect discourse in Sharvit (2008) and Eckardt (2014).[2]

This paper brings up a problem with this basic analysis that follows from the geometric content of pictures being so strong, and from the fact that human agents pick up relatively little information from their visual environments. The truth conditions delivered by the basic analysis are implausibly strong (i.e. implausibly hard to satisfy), in that the analysis delivers described situations where agents pick up more perceptual information than human agents normally pick up. This problem is explained in Section 2. Section 3 introduces modeling of the epistemic consequences of perceptual events using event alternatives, and defines a notion of normal looking events. Section 4 then applies these tools to the problematic data. Section 5 states a technical formulation. Section 5 sums up.

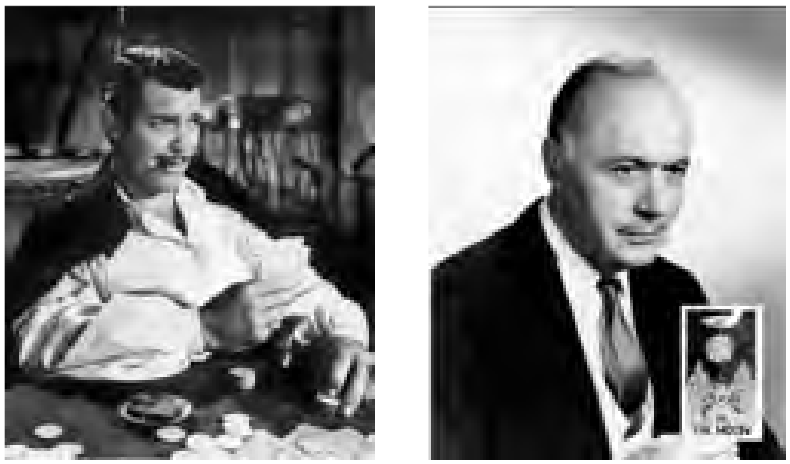
The discussion refers to the following constructed materials. Gable and Boyer are playing cards with an ordinary deck. In the sequence (6), Gable looks up (first panel) and sees Boyer holding

[2]In this literature, there is discussion of whether the embedding should be thought of as "syntactic" or "pragmatic", where pragmatic embedding would arise from something like a non-default interpretation rule. Here we pursue the strategy of using a compositionally interpreted syntactic level, comparable to a discourse representation as it figures in the account of natural language interpretation (Kamp and Reyle, 1993). That syntactic level is derived from surface pictorial narrative by shuffling in syntactic material such as the syntax that introduces discourse referents, but also by introducing operators that embed pictures that are top-level in the surface narrative. We don't take a position on the question whether such a level should be called pragmatic.

up an ace of spades (second panel). In the intensional version (7), Gable looks up (first panel) and hallucinating, sees Boyer holding up a moon tarot card (second panel).

(6)

(7)



In the second example, the non-veridical, intensional reading is enforced by the background assumption that Gable and Boyer are playing with an ordinary deck, so that Boyer is unlikely to hold up a tarot card.

A significant aspect of free perception sequences is that they are "character-centric", in the sense that their interpretation refers to a character who is depicted in the setup panel. Taking account of this requires machinery for indexing in pictorial narratives. Abusch (2012) introduced a syntax where discourse referents are introduced by areas in pictures, and referenced using numerical indices, with the index 1 referencing the most recently introduced discourse referent.[3] In the analysis from Abusch and Rooth (2017), these indices occur as arguments of operators that impose free perception readings. (8) is the LF of example (6). Here a is an area in the preceding picture, which picks out the depicted agent Gable.[4] V is an operator that imposes an extensional free perception reading. Semantically, it constrains the viewpoint for the second panel to coincide with the geometric visual viewpoint of the individual picked out by discourse referent 1 (namely Gable), so that the second panel shows what a described world

[3]See Abusch (2020) for a handbook presentation of the dynamic framework for pictorial narratives, and Rooth and Abusch (2018) and Rooth and Abusch (2019) for other applications of it.

[4]Depending on formulation, a could be a bounding box that encloses the projection of the character, or a point inside that projection.

looks like from the geometric visual perspective of the individual depicted in the first panel.

(8) LF of example (6)

 $a\ V(1)$ 

The LF (8) is extensional, in that a world in the interpretatiton of this formula has a local part that looks like the two panels from two nearby viewpoints. In such a world, a card player who looks like Boyer holds up an ace of spades within view of a card player who looks like Gable. This comes about because the second panel is a top-level conjunct in (8), and top-level panels always have extensional implications. An interpretation like this is not right for the hallucinating example (7), where we don't want to get the entailment that an agent who looks like Boyer holds up a moon tarot card in the base world of a described situation. This shows that in the LF of (7), the second panel must be embedded. (9) is a version of the syntax proposed in Abusch and Rooth (2017). *P* is an operator that imposes the intensional free perception reading. It has two arguments. The first is the discourse referent 1 for the agent depicted in the first panel. The second is a picture. *P* is effectily a covert verb "see", with $P(n,q)$ glossed as "agent *n* sees a view like picture *q*".[5]

(9) LF of example (7)

 $a\ P\left(1,\right.$  $\left.\right)$

Given that in (9) the second panel is the complement of a hidden covert verb "see", we might expect that verb to have a semantics modeled on analyses of intensional complementation in natural languages. This is exactly what is found in Abusch and Rooth (2017). The picture contributes a certain kind of intensional object, and the embedding verb is given a semantics that quantifies universally over epistemic alternatives.

Explaining this is deferred to Section 3. Here we note that the semantics entails that the agent picks up all of the information in the embedded picture, in the sense that in a verifying situation for (9), all of the agent's epistemic alternatives look like the second panel from the geometric perspecive of the counterpart of the agent.

This analysis treats intensional free perception readings in pictorial narratives as syntactically and semantically analogous to the intensional complementation seen in linguistics examples such as (10a). The indexed prounoun [he$_1$] corresponds to the discourse referent 1 which is the first argument of *P* in (9).

(10) a. He$_1$ saw a man who looked like Charles Boyer holding up a moon tarot card.

[5] Abusch and Rooth (2017) used a tree syntax, where *P* is a head with the picture as complement, and the discourse referent as subject.

b. He_1 hallucinated a man who looked like Charles Boyer holding up a moon tarot card.

2. Perceptual blindness

Perceptual blindness or inattentional blindness is the well-established phenomenon of human agents picking up relatively little information from their visual environments. 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). 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).

These facts interact with the semantics of free perception in the following way. As explained in detail in Section 3, an analysis of free perception sequences using Hintikka semantics has it that the agent picks up all of the information in the second panel of the sequence. The reason is that, in the formal model of the epistemic consequnces of perceptual events, an agent that perceives the ace scene has epistemic alternatives that look exactly like the ace picture from the geometric perscective of the counterpart of the agent. Equally, an agent who hallucinates the moon card has epistemic alternatives that look exactly like the moon picture from the geometric perspective of the counterpart of the agent. These results are undesirable, because human perceptual events do not have such strong epistemic consequences. Equivalently, human agents do not pick up as much information in perceptual situations as the primitive model has it.

Our strategy for solving this problem is a three-part one. First, a model construction is developed that takes perceptual blindness 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, using premise semantics. Finally, the normality condition is incorporated in the semantics for intensional free perception. This is accompanied by weakening universal quantification to existential quantification in the semantics of intensional embedding.

3. Event models and normal looking

The analysis uses a construction of possible worlds as event sequences, and a modeling of the epistemic consequences of perceptual events using event alternatives. Constructing possible worlds using events traces back to McCarthy's situation calculus (McCarthy, 1963; Reiter, 2001). Baltag et al. (1999) introduced event alternatives in epistemic semantics. Campbell and Rooth (2021) proposed a version of event-sequence and event-alternative possible worlds models that mathematically are guarded string models for an epistemic extension of Kleene Algebra with Tests (Kozen, 1997), and applied them as possible-worlds models in natural language semantics. We assume this constructive possible-worlds framework here.

To specify an event-based possible worlds model, one lists a primitive set of events, pre- and post-conditions for those events, and for each primitive event and each agent, an event-alternative relation. The latter is like an epistemic alternative relation in Kripke semantics for epistemic modality, except that it is a relation on events rather than worlds. The event relation is used to define an epistemic alternative relation on worlds for each agent, and thus to determine

event type	shorthand	precondition	gloss
	e_a	Gable faces 	Gable looking veridically at the ace scene
	e_m	Gable faces 	Gable looking veridically at the moon card scene
	e_a^m	Gable faces 	Gable facing the ace scene while hallucinating the moon card scene

event type	alternatives for Gable	alternatives for observer
		
		
		

Figure 1: Primitive treatment of the Gable-Boyer scenario in an event-sequence model.

a Kripke frame.

The primitive events are event *types*, in that they may "happen" in different worlds, and may happen multiple times in one world history. The three events listed in the left column of the top table of Figure 1 (along with others) are primitive events in a basic model for the Gable-Boyer scenario. The first event or

or e_a is an event type of Gable looking veridically at a scene just like the one shown in the ace picture. Epistemic consequences for Gable are captured by the set of *event alternatives* to this event for Gable. In this case, this is $\{e_a\}$, which is the unit set of the same event. This captures epistemic consequences of the event in the following way. Suppoose we are given a world w that satifies the preconditions of the e_a, a set wR of world alternatives to w for Gable, and a set of event alternatives e_aS to e_a for Gable.[6] When world w is incremented to a world we_a, the set

[6]Where w is a world and R is a relation on worlds, $wR = \{v|wRv\}$. Similarly where e is an event and S is a relation on events, $eS = \{e'|eRe'\}$. eS is the set of event alternatives to e.

of world alternatives $(we_a)R$ to we_a for Gable is defined to be

$$\{vd | wRv \wedge eSd \wedge v \text{ satisfies the preconditions of } d\}.$$

This is the set of worlds that can be formed by incrementing an alternative to w with and alternative to e_a, filtered by pre-conditions of events.[7] Since in this case the only option for event d is e_a, any world alternative to we_a for Gable is of the form ve_a. Since e_a has the pre-condition that Gable is facing a scene like the ace picture, in any world alternative Gable is facing a scene just like the ace picture. This captures the epistemic effect for Gable of picking up all of the information in the ace picture.[8]

To illustrate the modeling of epistemic semantics, in Figure 1 event alternatives are included for an observer who sees that Gable is looking, but does not see the card. For the observer, when e_a happens in the base world, e_a, e_m, or other events could happen in alternative worlds. This means that the observer has world alterantives of the form ve_a, ve_m, and ve for other events e that are looking actions of Gable. So after Gable looks at the ace scene, Gable has the information that an ace was held up, but the observer does not. The observer has the weaker information that Gable just looked.

The second event e_m listed in Figure 1 is the event of Gable looking veridically at the moon tarot scene. It is treated in the same way as e_a.

The third event e_a^m listed in Figure 1 is the event of Gable facing the ace, while hallucinating the moon card. An important idea is that alternatives to e_a^m are *ordinary* events e_m. This captures Gable picking up (incorrectly) the information that he is facing a moon card, since any alternative is of the form ve_m, and e_m has the precondition of facing the moon card. It also captures Gable not knowing he is hallucinating, and Gable believing that he believes he is facing a moon card.

This primitive analysis of the scenario in epistemic event semantics is subject to the criticism from Section 2, since it is designed to predict that in a free perception pictorial narrative, the agent depicted in the first panel gains all of the information in the semantic interpretation of the second panel. This is wrong because of attentional blindness. Fixing the problem requires changes both in the model construction, and in the syntax-semantics interface.

A clue to how to proceed comes from considering extensional free perception sequences. In the sequence repeated in (11), 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, he is subject to perceptual blindness, and would normally *not* pick up all of the information in the scene he is facing.

[7] The same definition applies for an arbitrary world we.

[8] See Baltag et al. (1999), Van Ditmarsch et al. (2007), and articles in Van Ditmarsch et al. (2015) for information about this way of modeling epistemic consequences of perceptual and communicative events.

(11)  $a\ V(1)$ 

Taking advantage of this requires a model where there are events of Gable looking while facing a scene like the ace picture that are weaker in their epistemic consequences than the event e_a defined in Figure 1.

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. 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 the Boyer's neckwear, e.g. necktie vs scarf; (iii) the configuration of the Boyer's eyes, e.g. open or closed. Let d_{xyz} be the event type of Gable looking at a scene with Boyer holding up card x, while wearing neckwear y, and with eyes in configuration z. In addition, looking events where Gable attends to a given feature should be distinguished from ones where he ignores that feature. This is recorded in the event types by bold-facing the features that Gable attends to. For instance, $d_{\mathbf{an}\text{o}}$ is the event type of Gable looking at a Boyer-like man holding up an ace of spades (letter 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 type $d_{\mathbf{a}\text{no}}$ is the event type 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.

In the basic construction from Figure 1, the set of event of alternatives to e_a was the unit set $\{e_a\}$, and this modeled the agent Gable picking up all the information in the scene he is facing. How should this be modified when e_a is extended to $d_{\mathbf{a}\text{no}}$ 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 $d_{\mathbf{an}\text{o}}$ for Gable is $\{d_{\mathbf{an}\text{o}}, d_{\mathbf{an}\text{c}}\}$, where there is variation in the eye-position features. Assuming there are two kinds of neckware n (necktie) and s (scarf), the set of event alternatives to $d_{\mathbf{a}\text{no}}$ for Gable is is $\{d_{\mathbf{an}\text{o}}, d_{\mathbf{an}\text{c}}, d_{\mathbf{as}\text{o}}, d_{\mathbf{as}\text{c}}\}$, where there is variation in the neckwear and eye configuration features, but not the card feature.

Events like this are a model of perceptual blindness. When such events occur in a base world, agents picks up some information from the scene they are facing, but not all. Suppose that in the card game, Boyer holds up the ace, and Gable attends to that, but not to Boyer's neckwear or to the configuration of Boyer's eyes, and in this way is "blind" to the necktie and the open eyes. This is modeled with the event $d_{\mathbf{a}\text{no}}$ transpiring in the base world. Weakened epistemic consequences are modeled by Gable having world alternatives that finish with any of the

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

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

events $\{d_{\mathbf{a}\text{no}}, d_{\mathbf{a}\text{nc}}, d_{\mathbf{a}\text{so}}$, or $d_{\mathbf{a}\text{sc}}\}$. Figure 2 lists some events of veridical looking in the more elaborate construction of the Gable-Boyer scenario.

Let us return to the narrative snippet (11). It 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 "the agent 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 neckware, he would normally attend to the neckware.

Further, what is attended to in a scene depends on the visual salience of the objects that are depicted. We will not build this into our toy constructions of events, but we could. Since stating a construction of events that is really faithful to what is known about human perception 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 (12). (12a) has already been located in the semantics of $V(1)$. Abusch and Rooth (2017) suggested that (12b) should be treated as accommodated, but since there is an operator V in the LF that is the syntactic correlate of extensional free perception, it might as well be treated as an entailment of this operator. (12c) needs to be formalized, and it needs to be located somewhere in the semantic

or pragmatic analysis.

(12) a. The agent picked out by discourse referent x is facing a scene like the second panel.
b. That agent does a looking action e.
c. e is a looking action that the agent could normally take while facing that scene.

The constraint in (12c) 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 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 that card, Gable will see that it is the ace of spades.

Kratzer (1981) described an application to circumstantial modality of premise semantics for modality, which is a general framework for the semantics of modals that was proposed in Kratzer (1978). Premise semantics uses a set of propositions called an ordering source to compare the normality or "closeness" of different options.[9] Options are optimized by maximizing the set of ordering propostions that are true.

Suppose we are given an event type e, a possible world w, and an ordering source O. We want to use O to assess whether we is a normal evolution of world w. This is done by comparing we to competitors we' for optimality.

In premise 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,

$$\{p|u \in p \wedge p \in O\} \supset \{p|v \in p \wedge p \in O\}.$$

If the relation above holds, u is more optimal than v in that moving from v to u adds propositions from O that are true.

Definition (13) defines a world as normal if there are no competitor worlds that are strictly more optimal.

(13) Normal worlds
Let u be a world, let A be a set of worlds, and O be an ordering source. Then u is normal relative to A and O, written $\mathcal{N}(u,A,O)$, if and only if there is no world in A that is strictly more optimal than u relative to O.

Consider how to apply normality to a world of the form $wd_{\mathbf{anc}}$, 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 L be the set of event types that are looking actions of Gable. This determines a set $\{we|e \in L \wedge w$ satisfies the preconditions of $e\}$ of worlds where Gable looks in one way or another at the same scene. This is used as the set A of alternative worlds in definition (13).

[9] More accurately, the ordering source parameter is a function from possible worlds to sets of propositions. This becomes relevant when modals are embedded.

Accordingly world $wd_{\mathbf{a}\text{nc}}$ involves normal looking if there is no alternative *we* where *e* is a looking action of Gable that is strictly more optimal according to the ordering source than $wd_{\mathbf{a}\text{nc}}$.

So far in this section, the model construction was enriched to model perceptual blindness, by including various events of an agent looking at a given scene, which differ in what features the agent attends to. The latter dimension has epistemic consequences, because the features that the agent attends to are held constant in alternatives, while features that the agent does not attend to may vary. The construction of worlds and their alternatives was cast in an event-sequence construction of possible worlds. Second, we characterized mathematically the notion of an act of looking at a given scene being a normal act of looking, relative to an ordering source.

A completely concrete analysis of the Gable-Boyer scenario would require constructing a specific ordering source. Suffice it to say that we want an ordering source that models human visual attention, so that for instance prominent depicted objects are attended to. The ordering source could be purpose-dependent, so that if Gable and Boyer are playing cards, Gable would attend to the ace, and if they are involved in a consumer panel on neckwear, he would attend to the necktie.

Should a normality entailment be included in the semantics of the free perception sequence (11)? Technically, it should be possible to build a normality condition into the semantics of V, which is the covert seeing predicate in extensional free perception. Including such a condition accounts for the fact that a reader of a comic 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 the author of the comic or the creator of the film intended for readers or viewers to draw this conclusion.

Normality assumptions are systematic in the interpretation of linguistic and pictorial narratives, though. In (14a), one automaticilly assumes that the skeleton key was used to open the door. In (14b), one assumes that the finger that was broken was Jack's. In a narrative where it has been established that a missing document is in the back pocket of a character's purse, (14c) is understood to imply that the character found the document. We think that in narratives, this kind of implication 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, it is not necessary or desirable to write it into the semantics of V.

(14) a. She took out the skeleton key and opened the door.
b. Jack broke a finger while catching a frisbee.
c. Finally she looked into the back pocket of her purse.

If we do want to write a normality entailment into the semantics of V, in our technical framework it is actually necessary to re-order the syntax. A narrative is evaluated relative to a world and a viewpoint, where the viewpoint is the geometric viewpoint for the last picture in the narrative. Since we want to say that the viewpoint for the second picture in the free perception sequence (e.g. the ace picture) is the geometric visual viewpoint of the agent depicted in the first picture, we need to have access to the viewpoint for the second picture. This motivates

switching the syntax to (15). With this, the strong semantics of V is stated in (16).

(15)  a  $V(1)$

(16) $V(k)$ is true with respect to a world u, tuple of witnesses for discourse referents $\mathcal{O}$, and viewpoint v iff
(i) u is of the form we, where e is a looking action by $\mathcal{O}[k]$, and
(ii) v is the geometric visual viewpoint of $\mathcal{O}[k]$ in u.

4. Normality in intensional free perception

The problem identified in Section 2 is that an analysis of intensional free perception using universal quantification over viewpoint-centered worlds satisfying the picture produces truth conditions that are overly strong, i.e. overly hard to satisfy. Such an analysis requires the agent to hallucinate all of the information in the embedded picture. This section first extends the model construction to include events of hallucination. Then the problem of overly strong truth conditions is addressed by weakening quantificational force in the semantics of embedding, and inserting a normality condition in the semantics of intensional free perception.

In the model construction from the start of Section 3, an event of Gable hallucinating the moon card while facing the ace in the base world was characterized as an atomic event e_a^m. The event alternatives as defined in Figure 1 were events e_m of Gable looking veridically at the moon card. This construction has to be modified, in view of the treatment of perceptual blindness in Section 3. The core idea will remain the same: event alternatives to the hallucinating events are events of veridical looking that correspond to the information that is hallucinated.

Consider a base-world event m of Gable hallucinating the moon card, while not hallucinating anything specific about the neckwear or the eyes. What are the event alternatives to m for Gable? Ordinary veridical events of Gable viewing the moon card are of the form $\mathrm{d}_{\mathbf{m},\text{-},\text{-}}$, 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 alternatives shown in (17).

Along the same lines, let m' be an event of Gable hallucinating the moon card and a necktie on Boyer's neck, without hallucinating anything specific about the eyes. This leads to the two alternatives shown in the bottom line of (17).

(17)

event	alternatives	event in notation defined below
m	$\{d_{\mathbf{m}\text{no}}, d_{\mathbf{m}\text{nc}}, d_{\mathbf{m}\text{so}}, d_{\mathbf{m}\text{sc}}\}$	$h^{m,-,-}$
m'	$\{d_{\mathbf{mn}\text{o}}, d_{\mathbf{mn}\text{c}}\}$	$h^{m,n,-}$

For a given choice of card, neckware, and eye configuration, the binary attention features generate eight events. In (17) it is stipulated that in the alternatives, the agent attends to the features that are hallucinated. That is, in the alternatives to m, Gable attends to the card features and not the other features, and in the alternatives to m', Gable attends to the card and neckware features, and not the eye state feature.[10]

[10] Here we make choices that result in straighforward epistemic interpretations for the events.

Generalizing this, we form event types for hallucinating events of the form $h^{x,y,z}$, 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 $h^{x,y,z}$ 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 $z = c$. Here the attention features are written as bit vector. For instance d_{mno100} is the event of Gable looking at a configuration with moon card, a necktie, and open eyes, while attending only to the card.

Section 3 referred to the hallucinating event

e 

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 event types 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 $h^{x,y,z}$ into event types $h^{x,y,z}_{x'',y'',z''}$ that are parameterized in the subscript by the scene that Gable is facing in the base world. The event alternatives are as before.

This model construction is an encoding of perceptual events and their alternatives that takes attentional blindness into account, and which allows for 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. 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 have the specified values, and are attended to.

Now we are ready to reformulate the semantics of the intensional free perception narrative repeated in (18). The earlier model construction and syntax-semantics interface had the consequence that a world *we* that satisfies (18) had epistemic alternatives $w'e'$ for Gable where Gable is facing a scene just like the moon picture, and where e' is a looking action by Gable that has the precondition that Gable is facing a scene just like the moon picture. This is problematic, because it has Gable gaining more information when he hallucinates than when he looks veridically.

(18)  $a\ P\left(1,\right.$  $\left.\right)$

In the result above, e' is a looking action by Gable, the pre-conditions of which entail that

Gable is facing a scene just like the moon picture. Our idea is to weaken this to e' being a looking action that Gable could normally take when facing a scene just like the moon picture. For instance, e' could be $d_{\mathbf{m}\text{no}}$, the action where Gable faces Boyer holding up a moon tarot card, wearing a necktie and with his eyes open, and Gable attends only to the card. This results in a larger set of epistemic alternatives for Gable, and so weakens the semantics of (18).

Definining the semantics of formula (18) can be posed as the problem of defining whether the formula is true or false in an arbitrary world of the form *we*. (19) lists different parts of the truth condition, and their sources. The first one says that world *we* looks like the first picture. This comes from the general semantics of pictorial narratives. The second one says that event e is a looking action by the agent in *we* that is picked out by the discourse referent a.[11] The third one is a placeholder for the intensional content of the predication introduced by P.

(19) Conditions for the truth of (18) in world *we*,
where x is an individual picked out in world *we* by discourse referent a, and R_x is the world alternative relation for x.

condition	source
we projects to p_1	Basic pictorial semantics
e is a looking action by x	P
Condition on $(we)R_x$ and $[\![p_2]\!]$	P

(20) lists some choices for the event e in *we*, and desired truth values. Formula (18) is false in *wc*, because c is an event of Gable coughing, rather than a looking event. Event h^{m--}_{ano} is the base event of Gable hallucinating a moon card. Formula (18) should come out true in a world of form wh^{m--}_{ano}. Event h^{-s-}_{ano} is a base event of Gable hallucinating a scarf. In a way that needs to be made precise, this event does not have epistemic consequences that are compatible with the embedded picture, and formula (18) should come out false in a world ending with this event. h^{-n-}_{aso} is the base event of Gable hallucinating a necktie in a situation where he faces an ace, scarf, and open eyes. For a more subtle reason, (18) will come out false in a world ending with this event.

(20)

event	desired truth value	description
c	false	Gable coughing
h^{m--}_{ano}	true	Gable hallucinating a moon card
h^{-s-}_{ano}	false	Gable hallucinating a scarf
h^{-n-}_{aso}	false	Gable hallucinates a necktie in a situation where he faces an ace, scarf, and open eyes

The intensional condition should be spelled out with reference to the semantic value $[\![p_2]\!]$ of the second picture, and to the world alternatives to *we* and/or the event alternatives to e. For the base events h^{m--}_{ano} h^{-s-}_{ano}, and h^{-n-}_{aso}, (21) lists the event alternatives for Gable in the toy model that was defined above.

[11] See below how this is stated using the technical mechanics of discourse referents.

(21)

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

Looking at the first case, updating the world alternatives to the world wh^{m--}_{ano}, where Gable has just hallucinated the moon card, results in worlds of the form $w_1 d_{\mathbf{m}\text{no}}$, $w_2 d_{\mathbf{m}\text{nc}}$, $w_3 d_{\mathbf{m}\text{so}}$, $w_4 d_{\mathbf{m}\text{sc}}$, 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(Hintikka, 1962).[12] As already discussed, insisting that all of the world alternatives satisfy this constraint will not work. For instance, in the toy model, in a world alternative $w_4 d_{\mathbf{m}\text{sc}}$, Gable is facing a scene where Boyer is holding up a moon card while wearing a scarf and having closed eyes. This information is imposed by the preconditions of event $d_{\mathbf{m}\text{sc}}$, and the information is inconsistent with the content of picture p_2, because in that picture, although Boyer is holding up a moon card, he is wearing a necktie and has open eyes.

A solution to this dilemma is to replace universal quantification with existential quantification: there is *some* world alternative to *we* that supports the content of the embedded picture. This seems unacceptably weak. Our idea is to fix this by including a normality condition: there is *some* world alternative $w'e'$ that satisfies the embedded picture (i.e. where $w'e'$ looks like p_2 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 $w'e'$. The notion of normality is the one which was introduced above with reference to veridical free perception. The witness for the existential condition in this case is $w_1 d_{\mathbf{m}\text{no}}$. This is a world where Gable is facing the moon card scene, and it ends with a veridical looking action where Gable attends to the card. Competitor worlds are formed by incrementing w_1 with looking actions by Gable (call this set $L(w_1)$). The ordering source O is assumed to be one according to which $d_{\mathbf{m}\text{no}}$ is a normal looking action for Gable when facing the moon card scene. Then the normality condition

$$\mathscr{N}(w_1 d_{\mathbf{m}\text{no}}, L(w_1), O)$$

is satisfied.

Consider on the other hand the base world of the form h^{-s-}_{ano}, where Gable has just hallucinated a scarf. It has alternatives of the form $w_1 d_{\text{a}\mathbf{s}\text{o}}$, $w_2 d_{\text{a}\mathbf{s}\text{c}}$, $w_3 d_{\text{m}\mathbf{s}\text{o}}$, or $w_4 d_{\text{m}\mathbf{s}\text{c}}$, where Gable faces various scenes, and looks veridically while attending only to the scarf. None of these worlds supports the content of p_2, because in p_2 Boyer is wearing a necktie.

For a case where normality is relevant, consider a base world of the form wh^{-n-}_{aso}, where Gable faces a situation with an ace, scarf, and open eyes, and hallucinates a necktie. This results in alternatives of the form $w_1 d_{\text{a}\mathbf{n}\text{o}}$, $w_2 d_{\text{a}\mathbf{n}\text{c}}$, $w_3 d_{\text{m}\mathbf{n}\text{o}}$, or $w_4 d_{\text{m}\mathbf{n}\text{c}}$, where Gable faces various scenes, and looks veridically while attending to the necktie. This includes a world $w_3 d_{\text{m}\mathbf{n}\text{o}}$ where Gable is facing the moon card scene, but it does not satisfy the normality condition

[12] The formulation here is a shorthand, which would have to be replaced with a formulation given in terms of the semantic content of p_2.

$$\mathscr{N}(w_3 d_{\mathbf{mno}}, L(w_3), O),$$

because normality according to O requires that Gable attend to the card when looking at the moon card scene. This illustrates that including normality strengthens the content of formula (18).

These truth values that were achieved above depend on the ordering source. If we think that formula (18) should entail that Gable ends up believing that Boyer is holding up a moon tarot card, in the analysis this comes from assuming an ordering source where any normal event of Boyer looking at the moon card scene involves him attending to the card.

5. Compositional formulation

Formulating the analysis from the preceding section compositionally is straightforward, because everything can be packed into the semantics of the hidden seeing predicate P. Syntactically, this predicate combines with a discourse referent and a picture. The truth condition for formula of the form $P(k,q)$ is given in (22). The argument k is a discourse referent (numerical index), and q is a picture. In addition to a world u, the definition refers to a tuple of individuals $\mathscr{O}$ that is used to provide values for discourse referents, and a family $\{R_x\}$ of world-alternative relations for agents.

Condition (i) is a way of saying that the individual picked out by index k is an agent. Condition (ii) says that world u ends with a looking action by the agent. Condition (iii) is the intensional entailment. It is an existential condition on world alternatives $w'e'$ to we. Part (iii)a says that world $w'e'$ looks like picture q from the geometric perspective of the agent. Part (iii)b is the normality condition. It says that event e' in $w'e'$ is a normal looking action, relative to an ordering source that defines normal looking, and competitor worlds that are formed by incrementing w' with a looking action by the agent.

(22) $P(k,q)$ is true with respect to a world u, tuple of witnesses for discourse referents $\mathscr{O}$, and family of world alternative relations $\{R_x\}$ for agents iff
- (i) $R_{\mathscr{O}[k]}$ is defined ($\mathscr{O}[k]$ is an agent)
- (ii) u is of the form we, where e is a looking action by $\mathscr{O}[k]$
- (iii) There is a world $w'e'$ in the set of world alternatives $uR_{\mathscr{O}[k]}$ such that
  - a. $\langle w'e', v\rangle \in [\![q]\!]$,
    where v is the geometric visual viewpoint of $\mathscr{O}[k]$ in $w'e'$
  - b. $\mathscr{N}(w'e', A, O)$, where A is the set of worlds of the form $w'e''$,
    where e'' is a looking action by $\mathscr{O}[k]$.

6. Discussion

Based on a naturalistically inspired interpretation of the epistemic consequences of looking that takes into account perceptual blindness, we argued that an analysis of intensional pictorial free perception that uses universal quantification is incorrect. Our analysis weakens universal quantication over epistemic alternatives to existential quantification, and compensates for the weakness of this by introducing a normality condition.

This line of analysis is surprising, because it departs so radically from the standard Hintikka semantics of complementation using universal quantification. Nevertheless, our logical form

for non-veridical free perception syntactically involves an embedded picture (rather than a picture that is a top-level conjunct), and semantically it refers to the intensional content of the embedded picture, and deploys it in a condition on epistemic alternatives.

A striking property of the argument in the paper is that it refers as much to a model construction using event alternatives as to the syntax-semantics interface for embedding. This contrasts with typical reasoning about the syntax-semantics interface in linguistic semantics, which pretty much takes epistemic alternative relations as stipulated. Reasoning this way was crucial for us, because we relied on particular assumptions about epistemic-alternative relations in the constructive framework.

Another striking property of the analysis is that it combines reasoning about possible worlds semantics and epistemic alternatives with reasoning and assumptions about the effect on agents of their environment. The semantics for pictures though is a possible worlds one, not a cognitive one where the semantics of pictures is modeled in terms of their effect on agents (Peacocke, 1987).

The empirical materials here were limited to free perception in the strict sense. Intensional phenomena in comics and film are not confined to free perception, though, and in a larger class of examples, the same issues arise. In an episode of the series *Poldark*, the protagonist has met an old flame in a graveyard, and kissed her. Later in the episode, he imagines confessing to his wife. In the shot for the imagined confession, Poldark is shown. So this shot does not assume the geometric perspective of counterparts to Poldark, and in this way the sequence is not comparable to intensional free perception. (Also, Poldark is understood to be imagining the confession, not hallucinating it.) Since the shot is visually detailed, it has geometric informational content stronger than what, plausibly, Poldark is being described as imagining. (23) is a frame from the shot.

(23) Poldark: I met Elizabeth. For the first time in years, we talked.

We hope that such examples will be amenable to a similar treatment.

What are the natural language analogues to the phenomena discussed here? There is a certain analogy to wide scope readings for nominals in attitudinal complementation, where some syntactic and semantic material in a complement clause does not contribute to the attitudinal predication. But our analysis does not look at all like standard analyses of wide scope readings. Similarly for de re interpretation, in the sense where this is distinguished from wide scope quantification. An analogy that is perhaps better is to "at-issue" distinctions (Potts, 2003). We could say that deploying the ordering source gives some of the information in the embedded

picture a different, profiled status, which we can call being at issue.

References

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

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

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

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

Baltag, A., L. S. Moss, and S. Solecki (1999). The logic of public announcements, common knowledge, and private suspicions.

Campbell, E. and M. Rooth (2021). Epistemic semantics in guarded string models. In *Proceedings of the Society for Computation in Linguistics*.

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

Grimes, J. (1996). On the failure to detect changes in scenes across saccades.

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

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

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

Kozen, D. (1997). Kleene algebra with tests. *ACM Transactions on Programming Languages and Systems (TOPLAS) 19*(3), 427–443.

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

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

McCarthy, J. (1963). Situations, actions, and causal laws. Technical report, Stanford CS.

Peacocke, C. (1987). Depiction. *The Philosophical Review 96*(3), 383–410.

Potts, C. (2003). *The Logic of Conventional Implicatures*. Ph. D. thesis, University of California, Santa Cruz 2003.

Reiter, R. (2001). *Knowledge in Action: Logical foundations for specifying and implementing dynamical systems*. MIT press.

Rooth, M. and D. Abusch (2018). Picture descriptions and centered content. In *Proceedings of Sinn und Bedeutung*, Volume 21, pp. 1051–1064.

Rooth, M. and D. Abusch (2019). Indexing across media. In *Proceedings of 22nd Amsterdam Colloquium*. ILLC, University of Amsterdam.

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.

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.