The Formal Semantics of Free Perception in Pictorial Narratives

Dorit Abusch and Mats Rooth

Cornell University, Ithaca, NY, USA

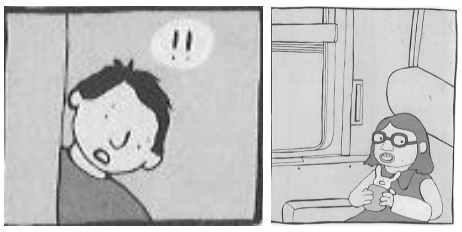
Abstract

This paper semantically analyzes "free perception" sequences in pictorial narratives such as comics, where one panel shows a character looking, and the next panel shows what they see. Pictorial contents are assumed to be viewpoint-centered propositions. A framework for the representation of pictorial narratives is used where indexing and embedding of certain panels is characterized by hidden operators. The resulting enriched pictorial narratives are interpreted in a dynamic framework. A possible worlds construction using action alternatives captures the epistemic effect of perceptual actions. Free perception sequences are implicitly anaphoric, as analyzed using cross-panel indexing. It is argued that some cases of free perception are truly intensional, and must involve embedding in the framework that is employed. Examples are drawn from comics and film.

1 Introduction

A common pattern in comics is a "free perception" sequence in which one panel shows a character looking, and the subsequent panel shows what is seen. The pair in (1a) is from S. Tan's *the Arrival*, showing a man looking down, and some enigmatic writing and graphics on the sidewalk.[1] It is understood that the second picture shows what the man sees. For another example, in Simone Lia's *Fluffy*, the character Michael has lost his rabbit Fluffy on a train. Searching, he looks into a cabin, and hallucinating, sees a girl eating a rabbit in a sandwich (see 1b). It is subsequently clarified that the girl was eating a kipferl, a kind of pastry.

(1) a. 

b. 

The same phenomenon is found in film. (2) shows three frames from *the Third Man*, showing a man looking off camera to his left, with the final frame showing what he sees.[2]

*Thanks to Ede Zimmermann for comments. A preliminary version of this work was presented at Göthe University, Frankfurt in summer 2017. Thanks to the audience for their reactions. The images in the paper that are quoted from comics and film are used for educational and critical purposes, and are property of their respective owners.

[1] *The Arrival* is entirely wordless, lacking captions, thought bubbles, and speech bubbles. Such works are of special interest in the study of pictorial narratives.

[2] Such "eyeline match" transitions are part of the system of film continuity editing. Cumming et. al. (2017) is a semantic study of aspects of this system.

(2) 

There are closely similar examples in natural language narratives (Brinton 1980). Frequently they consist of an eventive clause that describes someone looking, followed by a stative clause describing what is seen. See (3a-c). Sometimes the information that a character looks is accommodated, as in (3d).[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.

Current work on the semantics of pictures and pictorial narratives uses a possible-worlds model of information content (Greenberg 2011; Abusch 2012, 2016), based on the projective model of the semantic content of pictures (Hagen 1980). It is assumed here that a pictorial content is a viewpoint-centered proposition, modeled as a set of pairs of a world at a time and a geometric viewpoint (Rooth and Abusch 2017). A viewpoint is an oriented location in space, equivalent to the station point in the classical theory of perspective, or the location of an idealized camera. Functional notation is used for geometric projection, with $\pi(w, v, l, M) = p$ meaning that world-time w projects to picture p from viewpoint v. M and l are parameters for geometric projection.[4] Pictorial contents are obtained by inverting projection, $[\![p]\!]^{M,l} = \{\langle w, v\rangle | \pi(w, v, l, M) = p\}$.[5]

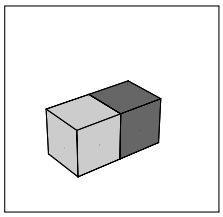
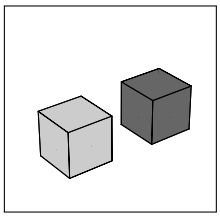
In order to model perceptual events and their epistemic properties, a construction of worlds as finite sequences of primitive events is assumed. Given a world w that satisfies the preconditions of an event a, wa is a world (at a time) where event a happened last. Perceptual events such as an agent looking come with event alternatives, and this is used in characterizing their epistemic properties. Thus we assume a construction of possible worlds as finite sequences of events, as in situation calculus (Reiter 2001), and a modeling of the epistemic consequences of events using Kripke relations on events, as in Baltag, Moss, and Solecki (1998).

Indexing across panels is significant in free perception sequences, because the agent about whom a free-perception picture gives visual-epistemic information is depicted in the previous panel. Characterizing the semantics of a free-perception panel involves reference to that agent, and this is a matter of indexing or anaphora across panels. Abusch (2012) introduced a syntactic approach to indices or discourse referents in pictorial narratives. Geometric points are interleaved with the narrative, and these points have the function of introducing and constraining model-theoretic values for discourse referents. Co-indexing is expressed with formal equalities. To illustrate, (5) is a short comic of two cubes moving apart, enriched with four discourse referents, and equalities between them. The notation is explained in a moment.

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

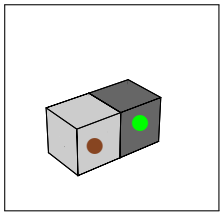
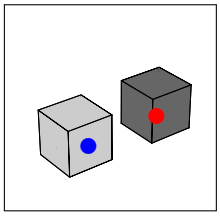
[4] l defines projection lines in terms of v, distinguishing for instance orthographic from perspectival projection. M is a marking rule that determines, for instance, that in (4), edges of geometric objects are marked in black.

[5] Abusch (to appear) is a survey of current work in this framework.

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

An enriched pictorial narrative provides information about a world, a viewpoint, and a sequence of individuals, with the latter functioning as witnesses for discourse referents. (5) illustrates the form of a semantic satisfaction clause, where a certain tuple satifies a certain enriched pictorial narrative to the right of the turnstile. w is a world-state, constructed as above; v is a viewpoint, interpreted as the viewpoint for the last picture, and (x_1, x_2, x_3, x_4) is a tuple of witnesses for discourse referents. (In (5), the colors and colored dots are not part of the formula.)

(5) $w, v, (x_1, x_2, x_3, x_4) \models$

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

Discourse referents are introduced with the interleaved geometric points. In (5), the point (0.7,0.5) is construed as a location in the preceding picture, and it introduces a discourse referent for the cube on the right in this picture (see the elements flagged in red). The point (0.3,0.3) introduces a discourse referent for the cube on the left in the last picture, flagged in blue. Similarly the points coming after the first picture introduce discourse referents for the cubes in that picture (flagged in green and brown). The semantics for discourse referents is random assignment, accompanied by a geometric constraint that locates objects in the model along a line determined by the current viewpoint and the geometric point specified in the discourse referent.[6] Formal equalities between natural numbers encode indexing across panels. A recency conventions is used: 1 is the most recently introduced discourse referent, 2 is the penultimately introduced discourse referent, and so forth. In (5), the equality 1=3 equates the dref for the cube on the right in the second picture with the dref for the cube on the right in the first picture. Similarly, 2=4 equates the drefs for the cubes on the left in the two pictures, which are flagged in blue and brown. The framework is comparable to a dynamic semantics for natural language where a discourse provides information about a world state and a list of individuals (Decker 2012).

The project for this paper is to use this toolkit to give a semantics for free perception in pictorial narratives. An important issue is the distinction between veridical free perception sequences such as (1a), where the free perception panel is construed as true of the base world timeline, and non-veridical ones such as (1b), where the base world timeline does not (or need not) satisfy the content of the free perception panel.

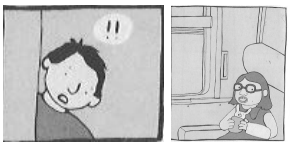
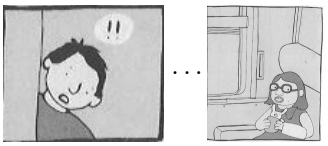
[6] See Abusch (to appear) for the details. Making it possible to state the semantics of discourse referents in this way is the motivation for storing the viewpoint for the last picture in the satisfying tuple.

2 LFs for free perception

We use the notation (p, q) for a free perception sequence, where p is the setup picture showing an agent looking, and q is the panel showing what the agent sees. In analyzing such sequences, there is an interplay between hypothesized logical forms for the narratives, interpretive principles for those logical forms, and modeling of the semantics of perceptual acts. We pursue a strategy of adding syntax to the narrative, in order to allow it to be interpreted incrementally and compositionally. Section 1 already mentioned that free perception sequences involve implicit anaphora to an agent in the first panel: a discourse referent for that agent should be added after the first panel, and then the semantics of the second panel should refer to that discourse referent. So a general hypothesis about the form of free perception sequences is (6), where p is the setup picture showing an agent looging, d introduces a discourse referent for that agent, and the complex $\phi(q, 1)$ interprets the second picture q in a way that explicitly or implicitly gives information about the visual-epistemic state of the agent. $\phi(q, 1)$ could involve syntactic embedding of q, or the addition of some conjuncts in a top-level sequence where q is a dynamic conjunct.

(6) $p\ d\ \phi(q, 1)$

To start, consider tuples that satisfy a non-enriched version of Fluffy sequence, as in (7a). Given the basic semantics, for any world w and viewpoint v that satisfy the narrative, $\pi(w, v, l, M) = q$, i.e. w looks like the second picture from viewpoint v. Of course, when we understand that Michael hallucinates, base worlds that satisfy the narrative do not (or need not) look like the rabbit sandwich picture from any viewpoint. The same point carries over to narratives with interleaved conjuncts. Whatever conjuncts are inserted in the position of the dots in (7b), any world w that satisfies the enriched narrative in the way shown in (7a) must have a prefix that satisfies the sandwich picture from some viewpoint. In other words, any narrative of the form seen in (7b) with the sandwich picture as a top-level conjunct entails (roughly) that a girl is eating or has eaten a rabbit sandwhich. This is the consequence of top-level pictures being interpreted extensionally, as providing information about what the base world (the world in the tuple to the left of the turnstile) looks like from some viewpoints at some times. Turning this result around, non-veridical free-perception panels are not top-level conjuncts.

(7) a. $w, v, \mathcal{O} \models \phi$  b. $w, v, \mathcal{O} \models \phi$  ...

We deal with this conclusion by hypothesizing covert embedding of non-veridical free perception panels. The syntax in (8) is inspired by the syntax of clausal embedding in natural language. P is a covert verb (roughly, "see") that embeds the free perception panel as a complement, and has the index 1 as its covert subject. This index picks up the discourse referent for Michael that is introduced by d after the first panel. Given this syntax, it is the semantics of the phrase headed by P, rather than the sandwich picture, that places a constraint on the world variable to the left of the turnstile. This semantics is taken up in the next section. The syntactic proposal is fairly minimal, in that it gives access to the free-perception panel and the perceiving agent, and by embedding the free perception panel, it blocks an extensional interpretation.[7]

[7] The proposal is syntactic in the same way that the introduction of discourse referents and equalities between

(8) $w, v, \mathcal{O} \models \phi$  d 1 P 

What about free perception sequences that are understood veridically? In (1a) we understand that worlds that satisfy the narrative do look like the second panel from the visual perspective of the agent depicted in the first panel. And we understand that worlds consistent with *the Third Man* look like the third image in (2) from the perspective of the man depicted in the first two images. Should an embedding syntax as in (8) be used also for such cases? Or for them, should an extensional syntax be hypothesized? We develop both options.

The idea for an extensional analys of sequences such as (1) and (2) is that the free perception panel is a top-level conjunct, but with a particular geometric viewpoint enforced. In the satisfaction clause (9) v to the left of the turnstile memorizes the viewpoint for q (here p is the setup picture, d introduces a discourse referent for the agent in p, and q is the free-perception frame). The recursive semantics ensures that w looks like q from v. This viewpoint v is in principle unconstrained, but here is understood to be a geometric viewpoint determined by the agent 1, corresponding to the location of the eyes (or other visual system) of that agent. Accordingly we add a geometric predicate $V(x)$, which contributes the geometric constraint that the ambient viewpoint is the oriented location of x's visual system. When $V(1)$ is added to the right of the free perception panel as in (10), it enforces that the viewpoint for the free perception panel is the geometric visual viewpoint for agent 1. In this, both the panel q and the predication $V(1)$ are extensional.

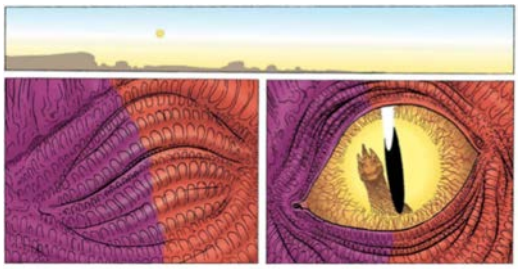
(9) $w, v, \mathcal{O} \models p\ d\ q$

(10) $w, v, \mathcal{O} \models p\ d\ q\ V(1)$

There are a couple of different panel types that are pragmatically similar to veridical free perception. (11a) is from Cece Bell's autobiographical *El Deafo*, and shows the heroine Cece and another character, Ginny. 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 quite 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 (11a) 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. But is arguably neutral about what information she picks up.

them is syntactic. In particular, it is the enriched narrative rather than the surface narrative that is interpreted compositionally. This way of proceeding is similar to what is seen in discourse representation theory (Kamp and Reyle 1993).

(11) a. 

b. 

(11b) is from Delgado's *the Age of Reptiles*. 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.

Consider for a moment what would be involved in an intensional syntax and interpretation for (11b). 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 genuine vision bubbles, as seen in (12a) from Bilal's *Cold Equator*. But such an analysis is otiose in the case of (11)b, because of the possibility of a straightforward extensional interpretation.

(12b) is from Tezuka's *Ode to Kirihito*. 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. Here again an extensional analysis is attractive.

(12) a. 

b. 

These three panel types (with sight lines, eye reflections, and over-the-shoulder viewpoint) are pragmatically 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 tend to 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 here too (assuming a switch in geometric viewpoint as enforced in (10)), the inferences that we make about what the agent sees are supported by the extensional content of the sequence.

3 Models for misperception and veridical perception

This section defines a model of perception in the event framework sketched in Section 1. The main idea is to model veridical perception and mis-perception using alternatives to perceptual events. The relation of alternativeness is like an accessibility relation in a Kripke model for knowledge and belief modalities, except that it operates at the level of events, rather than

worlds. This way of proceeding is based on Baltag, Moss and Solecki (1999).

We use the term $l(x, p)$ to represent the event of agent x looking veridically at a scene that projects to picture p from x's geometric perspective. This is an atomic event, which in the way reviewed in Section 1 figures in the construction of possible worlds. Such events have a role as event *types*, in that event $l(x, p)$ can occur in different world-time lines, or be repeated in a single world timeline. The properties of $l(x, p)$ are captured by its pre-conditions, and by its visual-epistemic alternatives for agent x.

Preconditions in situation calculus are used to capture the physics and metaphysics of the modal space. The elevator can go down only if it is above the ground floor. Block b can be placed on block a only if block a has a clear top surface. In the possible worlds model, world w can be incremented with event e to form world we if and only if the preconditions of e are true in w.[8] We think of $l(x, p)$ as a highly specific event of looking, which can happen only in worlds w where agent x is facing a scene that looks like p from the agent's geometric perspective. The position and orientation of agent x in w depends on the world history w—how x has moved in this history. The highly specific looking act $l(x, p)$ can happen in w only if that history is such that at the world/time w, x is facing a scene that looks like picture p. If this precondition is met, there is an incremented world $w \cdot l(x, p)$, where x has just performed an act of veridical looking.[9]

Epistemic properties of events are captured with a relation of event-alternatives. For a perceptual event e, taking the alternative-set for e to be the unit set e provides a modeling of veridical looking. Consider world a $w \cdot l(x, p)$, where $l(x, p)$ has just happened. Arguably any world of the form $u \cdot l(x, p)$ is consistent with the visual-epistemic information in the event $l(x, p)$ that just happened in $w \cdot l(x, p)$. In particular, because of the precondition, in u agent x is facing a p-like scene. If looking does not change the geometric facts, this is true also in $u \cdot l(x, p)$. Veridicality amounts to $w \cdot l(x, p)$ itself being a world of the form $u \cdot l(x, p)$, meaning that x is also facing a p-like scene in the base world. The agent is facing a p-like scene in both the base world, and any visual-epistemic alternatives for the agent. On top of this, the event $l(x, p)$ has just happened in the base world, and in any visual-epistemic world alternative. This is a kind of introspection condition on the source of the visual-epistemic information.

Using Q_x for the perceptual-alternative relation for agent x, these ideas are recorded in (13).

(13) Visual-epistemic event alternatives for $l(x, p)$
$Q_x(l(x, p)) = \{l(x, p)\}$

Visual-epistemic world alternatives determined by $l(x, p)$
$\bar{Q}_x(l(x, p)) = \{u \cdot l(x, p) | u \text{ satisfies the preconditions of } l(x, p)\}$

This account distinguishes the visual-epistemic content of the looking event from the epistemic state of the agent after looking. A world $v \cdot l(x, p)$ can be consistent with the perceptual information in the looking event that has just happened in $w \cdot l(x, p)$, but inconsistent with x's overall information in $w \cdot l(x, p)$. Let R_x be the epistemic alternative relation for agent x. (14) gives a principle in deduction format for updating R_x when a world w is extended with a perceptual action e of x to form $w \cdot e$. It amounts to what was seen before, but with the alternative $v \cdot e'$ required to be formed from a world v that is an epistemic alternative for x in w.

[8] See Reiter (2001) for a development of these concepts.

[9] Normally w can be extended in other ways, for instance with an axtion $s(x)$ of the agent stepping forward. So this is a branching-time model.

(14)
$$\frac{R_x(w, v) \quad Q_x(w \cdot e, v \cdot e')}{R_x(w \cdot e, v \cdot e')}$$

Discussions of free perception in language emphasize that it describes perceptual content, not epistemic state in the general sense (Kuroda 1976; Brinton 1980). Passage (3d) is understood to entail that Fred saw Jack coming across the street, not merely that he believed or knew he was. The same is true of pictorial free perception as analyzed using (13). $\bar{Q}_x(l(x, p))$ is a propositional content for the perceptual event $l(x, p)$, which is stated without reference to the epistemic state of the agent.

Veridical looking is characterized by visual-epistemic alternatives being similar to the base world in the way formalized in (13). In mis-perception, alternatives are not as similar to the base world. When Michael looks into the cabin, he sees a view q_r of a girl eating a rabbit sandwich. He believes he is engaged in veridical perception rather than mis-perception. This means his visual-epistemic world alternatives are of the form $v \cdot l(x, q_r)$, just as before. The difference is that the *base* world is not of this form. We introduce an additional basic looking action $m(x, p)$, thought of as an event of x looking at a scene which for x is p-like, but which is not (or is not necessarily) p-like in the base world.

(15) Visual-epistemic event alternatives for $m(x, p)$
$Q_x(m(x, p)) = \{l(x, p)\}$

Visual-epistemic world alternatives determined by $m(x, p)$
$\{v \cdot l(x, p) | v$ satisfies the preconditions of $l(x, p)\}$

For a simple idealized model, it is stipulated that events of the form $l(x, q)$ and $m(x, q)$ are the only looking events. A good setup panel and discourse referent for free perception is one which entails that the agent has just looked, i.e. that the last event that happened is either $l(x, q)$ or $m(x, q)$. or $w \cdot m(x, p)$. These are setup pictures where it "looks like" the agent picked out by the discourse referent is looking. We make the further assumption that actions of the form $l(x, p)$ are for the agent x alternatives only to looking actions. That is, if $l(x, p)$ is an element of $Q_x(e, l(x, p))$ then e is of the form $l(x, p')$ or $m(x, p')$.

Events $l(x, q)$ are used in scenarios of veridical looking, and events $m(x, q)$ are used in scenarios of mis-perception. Should it be assumed that events of the second kind are always erroneous, in the extensional sense that the base world does *not* look like q from x's geometric perspective? Consider a world w that looks like q from agent x's geometric perspective. World w satisfies the precondition of $l(x, q)$, and $w \cdot l(x, q)$ is a world where x has just looked veridically. If $w \cdot m(x, q)$ is also defined, then it is a formally different world which has the same visual-epistemic alternatives for x. So w branches into two worlds $w \cdot l(x, q)$ and $w \cdot m(x, q)$, that do not differ in properities that we want to model. This oddity is eliminated with a precondition for $m(x, q)$ that the world does not look like q from x's perspective (though the agent sees it as looking like q). We adopt this precondition for $m(x, q)$.[10]

[10] However, the other choice is also reasonable. If we think of $m(x, q)$ as x hallucinating a q-scene due to some specific effects in the low-level visual system or the cognitive system, it could be that x sees q due to those effects, but is accidentally right, in that x is facing a q scene in the base world. In this case, $l(x, p)$ happening should be distinguished from $m(x, p)$ happening, because only the first leads to knowledge. This comes up in Gettier scenarios.

4 Semantics for the LFs

This section interprets the LFs for free perception that where suggested in Section 2 in the event models from Section 3. (16) is the embedding LF, where q is embedded under P. The geometric point d sets up a discourse referent that can be referenced as $\mathcal{O}[1]$. ϕ is the part of the narrative preceding the free perception sequence.

(16) $w', v', \mathcal{O} \models \phi\ p\ d\ [1\ [P\ q]]$

Let w' be decomposed as $w \cdot e$, so that e is the event that just happened in w'. Where x is the agent $\mathcal{O}[1]$, $\overline{Q}_x(e)$ is the set of worlds that are perceptual alternatives to the event e that happens in the base world. Roughly, the semantics for the embedding construction should do a subset check between the visual-epistemic alternatives $\overline{Q}_x(e)$, and the content $[\![q]\!]^{M,l}$ of the embedded picture. Since content of the picture is viewpoint-centered, $\overline{Q}_x(e)$ needs to be adjusted to the viewpoint-centered proposition $\{\langle u', v'\rangle | u' \epsilon \overline{Q}_x(e) \wedge v' = \overline{V}(u', x)\}$, pairing the alternative world u' with the geometric viewpoint of x in u'. $\overline{V}$ is a function that maps a world and an agent to the geometric viewpoint of the agent in the world.[11] All of this leads to the semantics (17) for the embedding construction.

(17)
$$\frac{\begin{array}{c} w \cdot e, v, \mathcal{O} \models \phi \\ \mathcal{O}[n] = x \\ \{\langle u', v'\rangle | u' \epsilon \overline{Q}_x(e) \wedge v' = \overline{V}(u', x)\} \subseteq [\![q]\!]^{M,l} \end{array}}{w, v, \mathcal{O} \models \phi\ [n\ [P\ q]]}$$

A tricky question is what to do about the viewpoint in the conclusion. Normally a panel resets the viewpoint to the viewpoint from which the base world projects to the panel. In this case, since q is embedded, it is not projected in the base world, and there may be no viewpoint from which the base world projects to q. We have left the viewpoint constant.

On top of the truth conditions encoded in (17), it seems natural to say that $[n\ [P\ q]]$ presupposes that in w, $\mathcal{O}[n]$ is an agent with a visual system, and that e (the last event in $w \cdot e$) is a looking action by that agent. In the simple model construction where there are just two kinds of looking, $[n\ [P\ q]]$ presupposes that the base world finishes with either $l(x, q')$ or $m(x, q')$, for some q'.

(18) is the extensional option for the logical form of free perception. Herethere is nothing more to say about the semantics of q, since it is interpreted extensionally as placing a constraint on w and v. We just have to recall that $V(1)$ constrains v to be the geometric visual viewpoint of $\mathcal{O}[1]$, $v = \overline{V}(w, \mathcal{O}[1])$. This enforces that w looks like q from the geometric visual viewpoint of agent $\mathcal{O}[1]$.

(18) $w, v, \mathcal{O} \models p\ d\ q\ V(1)$

Section 2 finished with the question whether apparently veridical free perception sequences should be analyzed with the embedding LF (16), or with an LF where the free perception panel is in an extensional position as in (18). These options come out as symmetric in one dimension. The embedding LF expressed that things look like q for the agent, as expressed by the agent's visual-epistemic alternatives being of the form $u \cdot l(x, q)$. It presupposes that the agent is looking in the base world, but the base world could be either of the form $w'' \cdot l(x, q)$, with the agent facing a q scene in the base world, or of the form $w'' \cdot m(x, q')$, with q' not equal to q and the

[11] As Ede Zimmermann pointed out to us, it would be nice at this juncture if the alternatives were agent-centered worlds, rather than worlds. Then it would not be necessary to identify the agent across worlds.

agent facing some other kind of scene in the base world. The extensional LF (18) entails that w looks like q from viewpoint v. If w finishes with a looking event by x, it could finish either with $l(x, q)$, or with $m(x, q')$, with $q' \neq q$. Thus the embedding LF indicates what the agent's visual alternatives look like, and is neutral about what the base world is like. The extensional LF indicates what the base world looks like, and is neutral about what the agent's visual-epistemic alternatives look like. Resolving this issue requires further investication of what the entailments of examples such as (1a) should be.

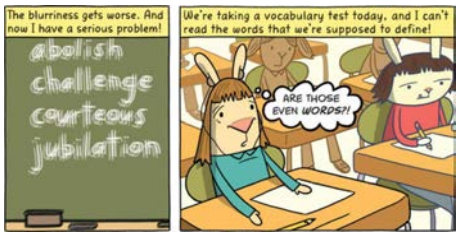
5 Discussion

This paper has developed LFs and a semantic analyses for two varieties of free perception sequences in pictorial narratives, veridical ones and non-veridical ones. The semantics used situation calculus models where worlds are constructed as sequences of events. Perceptual information was modeled using event alternatives. A dynamic system of interpretation was used to take account of the fact that free perception is implicitly anaphoric.

While there is not space to talk about it, a goal in this enterprise is to develop connections and contrasts between phenomena in pictorial narratives, and analogous phenomena in linguistic narratives and current theoretical conceptions of them. Current work on free indirect discourse, such as Sharvit (2008), Eckardt (2015), and Hinterwimmer (to appear) is immediately relevant. This is mainly concerned with a broader category of free indirect discourse. But many of the data discussed by Hinterwimmer can be considered examples of linguistic free perception, see (19).[12] A good way to proceed here would be to analyze linguistic free perception using the formal tools that were used in this paper, and compare results.

(19) The T-Rex hesitated. Maybe the little dinosaurs had hidden themselves in the cave on his left. When Billy looked up in his hiding place a few seconds later, a T-Rex bent down to the entrance of the cave and squinted into the dark.

The handful of classes of examples discussed here do not exhaust the phenomena of pictorial free perception. We mention without comment a couple of cases that we conjecture require a different analysis. In one passage of Bell's *El Deafo*, Ceci has blurry vision. (21a) is a free perception sequence, showing her view of a blurred blackboard.[13] In the film *the Terminator*, the Terminator has an infrared visual system, and views from its perspective are rendered as in (21b), using a red palette.

(20) a. 

b. 

Perceptual phenomena can be rendered in bubbles. In a passage in *El Deafo*, Ceci has obtained a hearing aid, and has gained hyper-acuity to sound. In (21), she hears a teacher in the bathroom. Here there are issues of a disjuncture between auditory and visual information. While the bubble structure seems to indicate embedding, semantically Ceci's auditory

[12] At this writing, Hinterwimmer's work is available to us as a handout.
[13] The sequence is inverted, something that is possible also for veridical free-perception sequences.

information could not be strong engough to entail the visual information in the picture.

(21)

We could continue for quite a while the list of examples that should fall under an account of depiction of perception in pictorial narratives, but are not covered by what has been said here. We hope that what we have proposed is a good starting point.

References

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

Abusch, D. (to appear). Possible worlds semantics for pictures. In Lisa Matthewson, Cécile Meier, Hotze Rullmann, Thomas Ede Zimmermann (eds.), *Blackwell Companion to Semantics*.

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. Morgan Kaufmann Publishers Inc.

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

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

Dekker, P. (2012). *Dynamic Semantics*. Springer.

Eckardt, R. (2015). *The Semantics of Free Indirect Discourse*. Brill.

Greenberg, G. J. (2011). *The Semiotic Spectrum*. PhD dissertation, Rutgers.

Hagen, M. A. (1980). *The Perception of Pictures. 1. Alberti's Window: the projective model of pictorial information*. Academic Press.

Hinterwimmer, S. (to appear). Two kinds of perspective taking in narrative texts. In *SALT 27*.

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

Kuroda, S. (1994). Reflections on the foundations of narrative theory from a linguistic point of view. In *Noam Chomsky: From artificial intelligence to theology.(2 v.)*, pp. 779. Taylor & Francis.

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

Rooth, M. and D. Abusch (2017). Picture descriptions and centered content. In *Sinn und Bedeutung 21*.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy 31*(3), 353–395.