Shifting perspectives in pictorial narratives[1]

Emar MAIER — *University of Groningen*
Sofia BIMPIKOU — *University of Groningen*

Abstract. We propose an extension of DRT for analyzing pictorial narratives. We test drive our PicDRT framework by analyzing the way authors represent characters' mental states and perception in comics. Our investigation goes beyond Abusch and Rooth (2017) in handling not just free perception sequences, but also a form of apparent perspective blending somewhat reminiscent of free indirect discourse.

Keywords: pictorial semantics, projection, visual narrative, comics, perspective, perception and attitude reports, free indirect discourse, DRT.

1. Introduction

Semantics is the study of meaning, where meaning is typically equated with truth conditions, in turn equated with sets of possible worlds. Semanticists study the truth-conditional contributions of various linguistic constructions (belief ascriptions, the German particle *ja*, etc.). But there are many non-linguistic human artifacts and activities that can be fruitfully analyzed as having truth-conditions and propositional content. Maps, gestures, traffic signs, paintings, mime performances, emoji, comics, ballet, music etc. Some of these can be highly effective means of communicating information, or of telling a vivid story. Some, like traffic signs, achieve their communicative function in a rather language-like way, with a more or less conventionalized, symbolic lexicon and some rules of composition. On the other side of the spectrum we have photography and mime, which convey information iconically, i.e. they depict what they are about by virtue of resemblance rather than mere convention.[2]

Recent advances in formal semantics, some under the header of 'super semantics', have aimed at capturing iconicity and depiction within the possible worlds framework that has proven so successful in linguistics (Schlenker, 2017, 2018).[3] In this paper we continue the pioneering work on the formal semantics of comics started by Dorit Abusch (2012, 2014). More specifically, since attitude and speech reports have proven such a driving force behind linguistic semantic theorizing ever since Frege (1892), we will here take a closer look at the ways attitudes are represented in comics. Our study thus adopts linguistic semantic tools to capture certain salient modes of attitude representation in pictures and comics, but in the end it turns out that our investigation of pictorial viewpoint shifting also sheds new light on linguistic viewpoint shifting phenomena.

2. Picture semantics

The centerpiece of a modern linguistic semantics formalization is a recursive definition of truth for a language, relative to a possible world: $[\![\varphi]\!]^w = 1$ iff .... Once this recursive definition is

---

[2] See Giardino and Greenberg (2015), Greenberg (2013), Schlenker (2018) for more on iconicity and resemblance in formal semantics.

[3] However, see also Zimmermann (2016) for a critical look at the idea that pictures, like sentences, express possible worlds propositions.

in place, the content or proposition expressed by a sentence φ is then easily defined as the set of worlds where φ is true. Pictures, especially photos, do not obviously have compositional structure, but they are about something and we can capture that with the help of a definition of truth in a world. A picture is true with respect to a possible world w iff there is a point v in w from where w 'looks like' the picture.

Following Abusch (2015), Greenberg (2013) and many others we can make this semantics more precise in terms of a projection function. A projection function π is a recipe for collapsing a 3D scene into a 2D plane. Formally, π takes a world w and a viewpoint v, and returns a picture p: $\pi(w,v) = p$. There are many different kinds of projection functions: we have linear, curvilinear, and parallel projections, projections that create black-and-white line figures with shading, and projections that retain a full range of colors realistically. We can think of this variation in terms of additional projection parameters, together defining a certain drawing style. In other words, a picture may be true of a world (from a given viewpoint) under certain parameter settings but false under certain other parameter settings. We will assume that the context determines the right settings of these parameters and then ignore them in our notations below.

Summing up, we use the projection function π (with contextually set parameters for perspective type, edge-to-line-conversion type, colors, etc) to define when a given picture is true of a world, from a given viewpoint:

(1) $\quad \left[\!\left[\text{} \right]\!\right]^{v,w} = 1$ iff $\pi(w,v) = $

From there we define propositional content (as the set of worlds in which the picture is true from some viewpoint):

(2) $\quad \left[\!\left[\text{} \right]\!\right] = \left\{ w \,\middle|\, \exists v \, \left[\!\left[\text{} \right]\!\right]^{v,w} = 1 \right\}$

3. From propositions to stories

3.1. The Dynamic Turn

Already in the 1980's, semanticists started looking beyond classical propositions to model the interpretation of utterances in the context of larger discourse. In dynamic semantics, the meaning of a sentence is not primarily the proposition it expresses, but rather the way it affects the context, i.e. the information conveyed by the discourse up to that point and further background knowledge shared by the participants. One particularly successful dynamic framework is Discourse Representation Theory (DRT, Kamp 1981). Below we illustrate the DRT framework with a textbook example of dynamic interpretation in language (from Geurts 1999), before we introduce an extension to DRT we call PicDRT that is able to model the dynamics of multi-panel visual story-telling.

In DRT, information is represented in the form of a Discourse Representation Structure (DRS). In (3b) we have represented the information conveyed by (3a), viz. that there exist a policeman and a squirrel and the former chased the latter, in the standard box notation for DRS's.

(3) a. A policeman chased a squirrel.

x y
policeman(x) squirrel(y)
chase(x,y)

b.

To interpret the next utterance in the discourse, (4a), we first add its compositionally gener-
ated logical form to the current discourse representation, representing pronouns as discourse
referents that need to be resolved (=?).

(4) a. He caught it.

x y z w
policeman(x) squirrel(y)
chase(x,y)
catch(z,w)
$z =?$ $w =?$

b.

We then proceed to resolve all the unknowns, by finding suitable antecedents among the ac-
cessible discourse referents. In this case, z (representing the pronoun *he*) binds to x (the po-
liceman), w to y. We may then simplify by unifying the equated discourse referents to get the
eventual output DRS in (5b).

(5) a.

x y z w
policeman(x) squirrel(y)
chase(x,y)
catch(z,w)
$z = x$ $w = y$

b.

x y
policeman(x) squirrel(y)
chase(x,y)
catch(x,y)

In this way we end up with an updated DRS representing the information conveyed by the two-
sentence discourse. A next sentence could continue adding information about the policeman
and the squirrel, using pronouns or other definites to refer back to these old discourse refer-
ents (*it bit him*), or it could introduce new discourse referents, using indefinites (*a woman was
standing nearby*). This way of cashing out the fundamental distinction between definites (intro-
ducing discourse referents that want to be bound to previously established ones) and indefinites
(introducing new discourse referents) is one of the basic features of DRT. It is also here that
pictorial discourse differs somewhat from linguistic discourse.

3.2. Introducing PicDRT

We can tell Geurts's simple mini-story also without words.

(6)
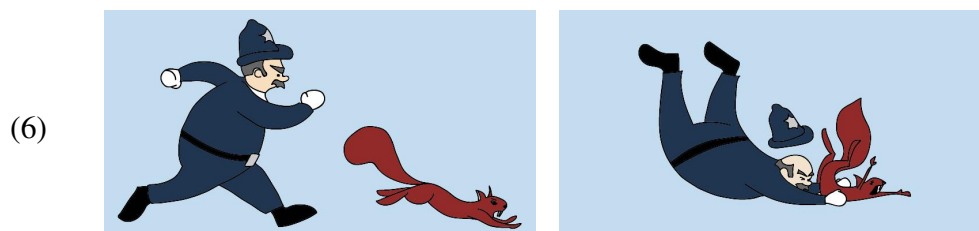
The juxtaposition of pictures in (6) is quite naturally interpreted as representing a temporal progression of two closely related events, the first picture shows a policeman chasing a squirrel, and the second shows him catching it. In this section we extend DRT with picture conditions and discourse referents to model that interpretation of sequential images (or comics, McCloud 1993, Cohn 2013b).

3.2.1. The PicDRT language

We'll extend the formal language of DRS's with picture conditions, consisting of a picture discourse referent (p_i) and an actual picture.[4] To capture the intuition that the policeman in the second picture refers to the same policeman as in the first, we need to impose some structure on pictures inside a PicDRS. We need to somehow identify discourse referents in the pictures (Abusch, 2012). We'll assume that the DRS construction algorithm manages to identify some regions of interest in a picture, viz. those regions that correspond to salient entities (like the policeman and the squirrel in (6)), and label those regions with fresh discourse referents. The first picture of the story in (6) will be represented in PicDRT as:

(7)
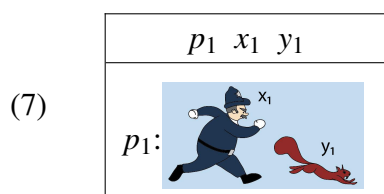
What this means is roughly that there is a situation or event[5] in the world that (from some viewpoint) looks like the picture, and in that situation there are two salient individuals, who look like the two regions labeled in the picture. We return to the model-theoretic/semantic details below, after we first illustrate the dynamics of the system.

3.2.2. PicDRT dynamics and pragmatics

The motivation for introducing discourse referents in pictures is to model the dynamics of storytelling with picture sequences. In the linguistic version of the story, (3)–(4), the second utterance could refer back to previously established discourse referents by using pronouns. As Abusch notes, there are no obvious pictorial analogues of pronouns (or definites or presupposition triggers more generally). There is nothing in the second picture that specifies that the depicted squirrel and policeman must be familiar from previous discourse rather than representing two new agents that just look similar. We'll assume, with Abusch, that each picture in a sequence introduces its own new discourse referents corresponding to salient regions, and it is left to pragmatics to determine whether some are to be treated as coreferential. Updating the representation in (7) with the second picture in (7) thus yields (8a). Pragmatic reasoning

---

[4]This makes a lot of sense if we think of DRS's as mental representations, which are quite naturally thought of as partly symbolic, partly iconic (visual/pictorial).

[5]We leave the tricky ontological details for another occasion (e.g. Should we use situation or event semantics? Should we label the event(s) with event discourse referents in the picture? Can an event even be depicted?)

based on assumptions of coherent storytelling will allow the reader to conclude that the second policeman is likely the same as the first, and similarly for the squirrel, which gives us the pragmatically strengthened representation in (8b).

(8)   a.

p_1 x_1 y_1 p_2 x_2 y_2
p_1: p_2:

b.

p_1 x_1 y_1 p_2 x_2 y_2
p_1: p_2:
$x_2 = x_1$ $y_2 = y_1$
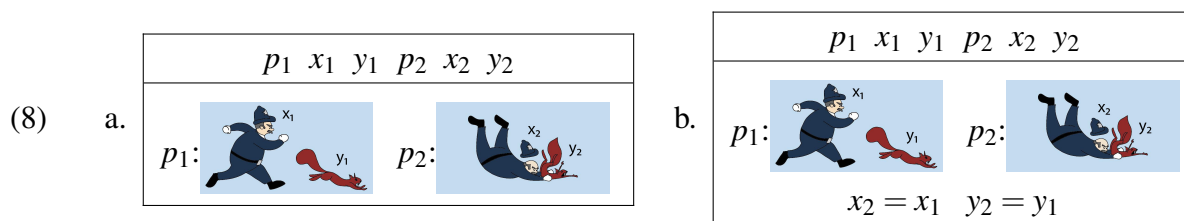
The pragmatically enriched PicDRS output in (8b) means roughly the same as the DRS output we derived for the linguistic story before: there's a situation that looks like the first picture, with two agents, looking like the two labeled regions, and there is another situation that looks like the second picture, with these same two agents, now looking like like the two labeled regions in the second picture. However, the ways in which we derived those outputs are importantly different. In the linguistic case, the coreference was encoded in part by the linguistic structures (pronouns) themselves, while in the pictorial case, it's purely pragmatic, i.e. it's ultimately up to the interpreter to interpret the second policeman as the same as in the first picture, or as a completely different new one. Similar points can be made about the temporal, aspectual, causal, and coherence relations that connect the two discourse units. For instance, the fact that the second sentence is interpreted as describing an event immediately following the first is partly determined by the choice of tense and aspect morphology on the verbs. In the pictorial version, there is no (obvious) analoguous morphology expressing temporal progression. Indeed, as McCloud (1993) already notes, juxtaposition of pictures in a comic may occasionally correspond to overlapping state descriptions, simultaneous shots from different viewpoints, or flashbacks and jumps in time, so we'd do well do leave this temporal ordering, again, to pragmatic strengthening. We will however assume that, by default, a picture to the right or below (in Western comics and picture books) another corresponds to a state of affair that comes later. We model this by adding a DRS condition of the form $p_1 < p_2$.

(9)

p_1 x_1 y_1 p_2 x_2 y_2
p_1: p_2:
$x_2 = x_1$ $y_2 = y_1$ $p_1 < p_2$
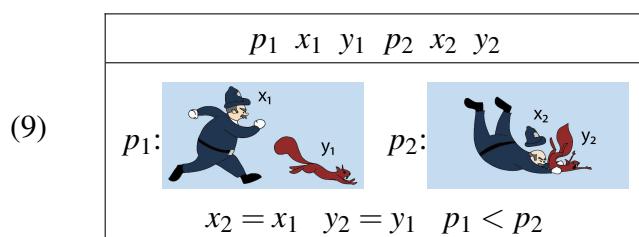
3.2.3. PicDRT semantics

In standard DRT we define when a DRS is true relative to a world w and a partial assignment or embedding f mapping some discourse referents to individuals in the domain of the model (as usual for any first-order language). For PicDRT we'll need to add a third parameter v providing viewpoints for the interpretation of the picture conditions (as described in section 2). Since pictures in a comic tend to represent the world from a variety of viewpoints (in space and time) we need our v to provide not one but a sequence of viewpoints, one for each picture condition. Formally, let's say v is a partial mapping from the set of pictorial discourse referents

$(\{p_1, p_2, p_3 \dots\})$ to points in space-time. Viewpoint functions then can be treated analogously to standard DRT's verifying embeddings, but for pictorial discourse referents.

Let's illustrate with our (pragmatically strengthened) PicDRS output, (9), from above. The starting point is that we use an assignment function f to verify the regular DRS conditions (relative to w), and a viewpoint function v to verify the pictorial conditions (also relative to w, using the projective definition of truth from section 2 above). Finally, since the regular, descriptive discourse referents also correspond to picture regions we must make sure that f and v are properly aligned (see (10c)).

(10)

$=1$ iff . . .

. . . iff there is a verifying embedding $f' \supset f$ with $Dom(f') = \{x_1, y_1, x_2, y_2\}$ and a viewpoint function $v' \supset v$ with $Dom(v') = \{p_1, p_2\}$ such that:

a. f' verifies the descriptive conditions:
  (i) $f'(x_2) = f'(x_1)$
  (ii) $f'(y_2) = f'(y_1)$

b. v' verifies the pictorial conditions:
  (i) $\pi(w, v'(p_1)) = $
  (ii) $\pi(w, v'(p_2)) = $
  (iii) $v'(p_1) < v'(p_2)$ (the viewpoint associated with p_1 temporally precedes that associated with p_2)

c. f' and v' are aligned:
  (i) $\pi(f'(x_1), v'(p_1)) = $ (the policeman-region in the picture is a projective (partial) depiction of the policeman represented by x_1, from the viewpoint associated with the picture)
  (ii) $\pi(f'(y_1), v'(p_1)) = $
  (iii) $\pi(f'(x_2), v'(p_2)) = $
  (iv) $\pi(f'(y_2), v'(p_2)) = $

Note that the extension of DRT semantics sketched here allows us to deal with recursive embedding of PicDRS's in complex conditions with intensional operators, which we'll encounter below. It also allows us to define the usual semantic notions of content, like the classical propositional content of a PicDRS:

(11) $[\![K]\!] = \{w \mid [\![K]\!]^{w,\emptyset,\emptyset} = 1\}$

4. Picturing speech and thought

In natural language semantics, from its Fregean beginnings in philosophy, attitude and speech reports have always played an important role. As a case study for our PicDRT framework, let's see if we can adequately describe the various attitude reporting strategies in pictorial narratives, starting with a few remarks about the obvious speech and thought bubbles, through free perception sequences, and ending with what we call blended panels.

4.1. A note about bubbles

One of the most recognizable features of contemporary comics are speech and thought bubbles. These devices are used to convey a character's utterances or inner thoughts, as is illustrated in (12) below. Usually, the utterances and thoughts are represented verbally, in written language. Such bubbles can be straightforwardly analyzed as the visual language analogue of quotation marks in language (Saraceni, 2003). Interestingly, comics also allow for the use of pictures inside thought (and, more rarely, speech) bubbles to represent an agent's thoughts or other mental states iconically.[6]

(12)   a.   b.
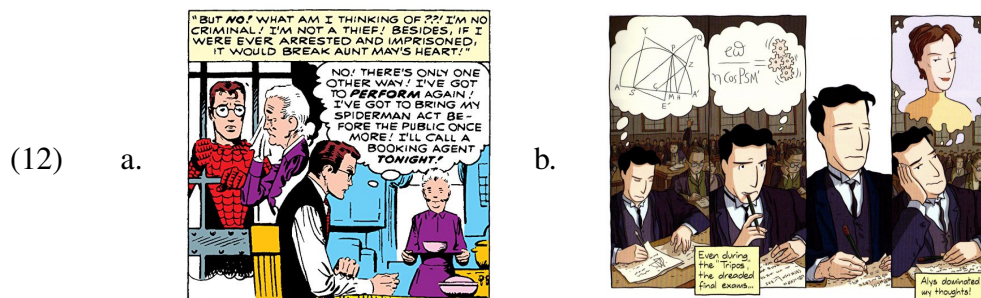
We choose to leave a detailed analysis of speech and thought bubbles as a form of quotation (or demonstration) for a future occasion. Instead we want to focus here on arguably more purely pictorial modes of representing mental states. For an extensive analysis on speech and thought bubbles in comics within a different kind of framework, see Cohn (2013a).

4.2. Free perception and viewpoint shift

Abusch and Rooth (2017) discuss the phenomenon of free perception sequences in wordless comics as a way to represent what a character is seeing. Typically, we have a panel showing a character looking, followed by a panel depicting what they see, as if through their eyes.[7]

---

[6](12)a: *The Amazing Spider-Man #1*, 1963, Marvel Comics; (12)b: *Logicomix: An epic search for truth*, 2015, Bloomsbury Publishing USA.

[7](13)a: *Hostage*, 2017, Vintage Publishing. (13)b: *Cyanide and Happiness*, 2018, `http://explosm.net/comics/4913/`.

(13) a.   b.

Following Abusch and Rooth (2017) we can analyze this kind of perception representation purely extensionally: the two pictures just describe the same (set of) world(s), but from two different viewpoints. Basically, interpreting the sequence involves fixing the viewpoint for the perception panels on to the location of the protagonist's eyes (at the looking time).

We can model this in PicDRT by introducing a special predicate 'view(x, p)' to capture the viewpoint shift. Note that the view-condition is, again, a pragmatic inference, drawn by the careful reader familiar with film and comic conventions like the setup panels focusing on the character's eyes.

(14)
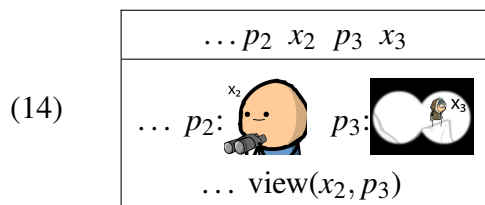
As stated, the semantics of this view-condition is purely extensional:

(15) $[\![\text{view}(x_2, p_3)]\!]^{f,v,w} = 1$ iff $v(p_3)$ corresponds to the location of the eyes of $f(x_2)$ (and the time of looking) in w

On this approach we're not really dealing with the representation of perception as a mental state.

The question arises whether there might be a linguistic analogue of this type of free perception representation. Abusch and Rooth (2017) refer to work by Brinton (1980) and Hinterwimmer (2017), who discuss the phenomenon of 'represented perception' in narratives (see also Banfield 1982 on 'representing non-reflective consciousness'). Two examples are given below:

(16) a. He looked at his mother. Her blue eyes were watching the cathedral quietly. (D.H. Lawrence *Sons and Lovers*, cited by Brinton 1980)
      b. Sara got up and went to the window. A crowd had gathered outside the public house. A man was being thrown out. There he came, staggering. (Virginia Woolf *The Years*, cited by Banfield 1982)

In (16a) the second sentence is most naturally interpreted as a description of what the subject of the first sentence was seeing. In (16b) the choice of the deictic *came* (rather than *went*) suggests that indeed the scene is described from the perspective of the salient viewing character, Sara.

4.3. Non-veridical free perception

Apart from evidently extensional viewpoint shifting – i.e. veridical free perception, in Abusch and Rooth (2017) terminology – there are also cases in which a narrative describes a fictional character's visions, dreams or hallucinations that are clearly not veridical. In (17) Rorschach is shown an inkblot which triggers a visual memory of a dead dog.

(17)

Unlike in the sequences in (13) above, the dead dog is just in the protagonist's mind and is not part of the actual scene. In (18), Bart Simpson is shown looking at an empty jar, followed by a representation of what he sees, a dead fairy in the jar.

(18)
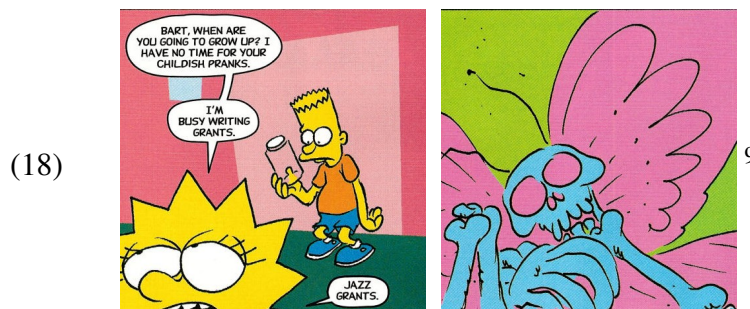
As in the previous cases of free perception, the second panel seems to depict the situation as viewed from Bart's geometric viewpoint, but instead of depicting the story world itself, it depicts the world according to Bart. Similarly, the last picture from the *Watchmen* sequence in (17) shows what Rorschach brings to his memory while looking at the drawing. In this sense, the picture is also a case of non-veridical perception as it does not represent the actual scene, but represents a scene from a different time that the protagonist brings to his mind.

In order to capture non-veridicality in our framework, we need an intensional operator here. Abusch and Rooth essentially posit such a hidden operator in the syntactic structure of the second picture. By contrast, we adopt a pragmatic approach and derive the insertion of a relevant operator on the basis of Eckardt's (2014) notion of a Cautious Update. To introduce this concept we take a little detour into the pragmatics of non-cooperative communication.

Linguists tend to assume idealized cooperative exchange situations in which the speaker intends to provide reliable information and the hearer trusts the speaker. In this case, asserting a proposition p means adding it to the common ground. This, Eckardt calls a Trust Update. However, a hearer may well distrust or disagree with the speaker. In such cases, the hearer may not accept the speaker's assertion that p and refrain from adding p to the common ground.

---

[8] *Watchmen*, 1987, DC Comics.
[9] *Bart Simpson's Treehouse of Horror #12*, 2006, Bongo Comics.

Instead, the hearer may accept something weaker, e.g. that the speaker believes that p. In Eckardt's terminology, we do a Cautious Update: instead of accepting and updating with p we update with $Bel_x p$.[10]

We can now view our non-veridical sequences as requiring the reader to perform a kind of Cautious Update, because a straightforward Trust Update will lead to an incoherent output where an inkblot suddenly changes into a dead dog, or a jar is empty the one moment and then contains a mythical creature the next. In PicDRT, we distinguish now two types of update. So far we've modeled the PicDRT Trust Update, which will always be the default interpretation strategy. Applied to the Simpsons comic, this standard update would lead to an incoherent output, even if we infer an extensional viewpoint shift condition, as shown in (19):

(19)
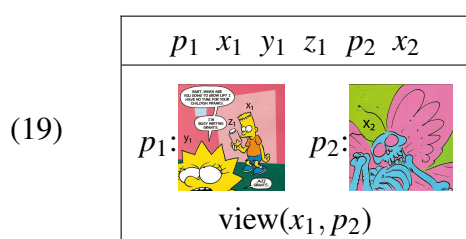
On this reading, Bart holds an empty jar and as soon as he looks at it a dead fairy appears. This is inconsistent with the rest of the story because it would predict that Lisa and others would be able to see the creature whenever Bart looks at it, which is clearly not the case. Whenever a Trust Update fails to get a coherent output we may resort to Cautious Update, which means that we add a suitable attitude operator to the information from the second picture. For these pictorial cases we'll assume a perceptual attitude operator, $PERC_x$, meaning something like 'x mentally perceives that …', (20b). We'll further assume that the referent for the perceiving agent x can be bound to any currently salient discourse referent.

(20)   a.
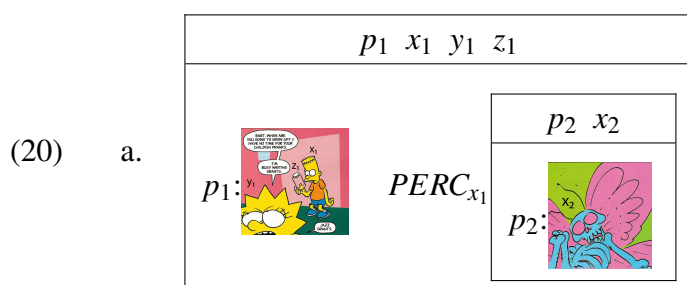
   b. $[\![PERC_{x_1} K]\!]^{f,v,w} = 1$ iff for all $\langle w', v' \rangle$ compatible with x_1's *de se* 'perceptual experience' (i.e. $v'(p_i)$ is the viewpoint from where $f(x_1)$ experiences world w' for all p_i in $U(K)$): $[\![K]\!]^{f,v',w'} = 1$

Note that by using a monstrous operator $PERC_x$, modeling *de se* perception, we account for both the free perception's visual viewpoint shift and the non-veridical, attitudinal embedding at once.

We conclude this discussion of free perception by again pointing out that there are seemingly

---

[10]Eckardt goes on to use this mechanism also for the interpretation of free indirect discourse. See Altshuler and Maier (2018) for another application of Cautious Update in DRT. See Asher and Lascarides (2013) for more on non-cooperative conversation, and Kamp (2016) for another way to handle updates from unreliable conversion partners in a mentalistic variant of DRT.

analogous phenomena in the linguistic domain. The following example, from *The Shining*, describes a scene where the hotel manager gives Jack, Wendy and Danny a tour of the hotel and Danny has a vision while they are admiring the presidential suite:[11]

(21)   Jack and Wendy were so absorbed in the view that they didn't look down at Danny, who was staring not out the window but at the red-and-white-striped silk wallpaper to the left, where a door opened into an interior bedroom. And his gasp, which had been mingled with theirs, had nothing to do with beauty. Great splashes of dried blood, flecked with tiny bits of grayish-white tissue, clotted the wallpaper. It made Danny feel sick.

The splashes of dried blood are part of Danny's psychic vision. It is clear from the rest of the story that there really is no blood to be seen there. The narrator describes the scene as mentally perceived by Danny. In our Eckardt-based approach, this means the reader is supposed to perform Cautious Update, interpreting the final passage as embedded under a mental perception operator like in (20b).

4.4. Perspective blending

We have dealt with sequences of pictures involving perspective shifting from one picture to another. However, in comics (as well as in other media like film) it is very common to depict a character and their subjective experience (faulty perception, imagination, dreams etc.) in a single image (or scene, in the case of film). In (22a), from the same *Simpsons* story as the above, Bart is depicted as holding the jar with the fairy inside, though it's still evidently only visible to him. In (22b) Batman's sidekick Robin is hit by Scarecrow's fear toxin and has to confront his own worst nightmares, which are depicted from various angles, but none of them from the perspective of Robin himself (as he himself is visible on almost every panel). In (22c) we first see Calvin with his friend Susie; followed by a depiction of that same scene in the way Calvin himself imagines it.[12] [13]

---

[11]Stephen King, *The Shining*, 1977.

[12]Zimmermann (2016) points out a similar case of blended, non-veridical perception in Ferdinand Bol's 1642 painting *Jacob's dream*. Another example is Antonio de Pereda's 1650 *The Knight's Dream* which depicts a sleeping knight next to the contents of his dream, an angel and various symbols of vanity.

[13](22)b: *Batman and Robin Eternal #2*, 2015, DC Comics; (22)c: *Calvin and Hobbes*, Andrews McMeel Publishing.

(22)   a.

      b.   c.

Because such images depict the perceiving, dreaming or hallucinating character (from an apparently neutral viewpoint), alongside the contents of their subjective experience, it seems that two perspectives are 'blended' in a single image. Such mixing of perspectives is reminiscent of the phenomenon of free indirect discourse in linguistic narrative, which is likewise characterized as blending two simultaneous perspectives or voices. Consider the example in (23).

(23)   She looked at the calendar. But... then the deadline was tomorrow! How was she supposed to fix that bloody paper in one day?

The second sentence in (23) describes what the protagonist is thinking, while looking at the calendar. On the one hand it thus represents the world from the protagonist's perspective, faithfully capturing her mood by means of exclamation, hesitation, and question marking, and by the use of indexicals like *tomorrow* reflecting the protagonist's perspective. On the other hand, the third person pronouns and past tenses reflect the neutral narrator's perspective.[14]

There are two main types of semantic approaches to free indirect discourse. The quotational approach treats it as a form of quotation, where pronouns and tenses are systematically unquoted (Maier, 2015, 2017). Bicontextual approaches introduce a second, shifted context parameter ('context of thought' or 'protagonist context') that takes care of the shifted interpretations of indexicals like *tomorrow* in (23) (Schlenker, 2004; Eckardt, 2014). Either way, the effect is that part of a free indirect discourse passage will be semantically interpreted relative to the neutral, narrator's perspective/context, while other parts get interpreted relative to the protagonist's perspective/context. So let's see if we can translate that insight to our pictorial blends.

Take another example, (24a), where we see the protagonist, Joe, and his hallucinations – his toys having come to life and surrounding him. Following the free indirect discourse approach, part of the picture should represent the actual state of affairs in the story, the part depicting Joe lying on the floor, while the rest of the picture represents Joe's subjective experience. Sticking

---

[14]For a short overview of different narratological theories about Free Indirect Discourse, see Bray (2007).

with a free indirect discourse approach thus entails that in interpreting (24a) we essentially split the picture in two parts, separating the experiencer, depicted from the neutral perspective, from the subjective experience itself, as in (24b).[15]

(24)   a.   b.

In this way, interpreting the original blended picture is reduced to splitting the perspectives (just like we separate the quotation/protagonist-oriented and unquotation/narrator-oriented parts of a free indirect discourse in our familiar semantic approaches to the linguistic phenomenon) and then interpreting the result essentially as a non-veridical free perception sequence.

On closer inspection this linguistically inspired approach gives the wrong result. Unlike in a free perception sequence, the viewpoint remains stable in (24b). The subjective experience part, where Joe is cut out, presents Joe's mental state, but not from his perspective. The *Calvin and Hobbes* passage in (22)c makes this even clearer. The blended panel presents the scene from the exact same neutral viewpoint as the previous panel. In addition, note that it would be impossible to split that picture into an experiencer, Calvin, and his experience, as his experience affects the way Calvin himself is depicted (as an astronaut).
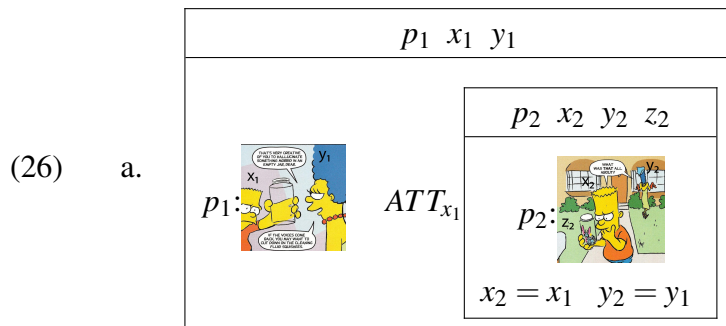
In sum, a blended picture presents a scene from a single viewpoint, the narrator's, not the protagonist's, and is thus not strictly speaking a case of 'dual voice' after all. Instead, we can think of the blended panels as the analogues of (nonmonstrous) indirect discourse and attitude reports in English.

(25)   Joe dreamed that he was surrounded by his toys

An indirect report like (25) shows no shifted indexicals that might indicate quotation or some other shift to a protagonist context. Everything in (25) is interpreted from the narrator's neutral perspective, all we have here is an intensional operator quantifying over Joe's dream worlds.

In PicDRT we propose that Cautious Update may introduce, instead of a monstrous, *de se* perception operator ($PERC_x$ from (20b)), also a non-monstrous, purely intensional attitude operator (ATT_x). For the blended Simpsons panel in (22)a this results in the output in (26):

---

[15](24)a: *Joe the Barbarian*, 2011, Vertigo Comics.

(26) a.

$$
\boxed{
\begin{array}{l}
p_1 \ \ x_1 \ \ y_1 \\[4pt]
\hline \\
p_1: \text{<comic panel>} \quad ATT_{x_1} \quad
\boxed{
\begin{array}{l}
p_2 \ \ x_2 \ \ y_2 \ \ z_2 \\[2pt]
\hline \\
p_2: \text{<comic panel>} \\[4pt]
\hline \\
x_2 = x_1 \quad y_2 = y_1
\end{array}
}
\end{array}
}
$$

b. $\llbracket ATT_{x_1} K \rrbracket^{f,v,w} = 1$ iff for all w' compatible with $f(x_1)$'s attitudinal state in w: $\llbracket K \rrbracket^{f,v,w'} = 1$

5. Conclusion

Pictures, like sentences, have truth-conditional content. And sequences of pictures, like sequences of sentences, can be used to tell stories. We introduce PicDRT, a simple extension of standard DRT to analyze pictorial storytelling in a dynamic semantic setting.

The current paper presents a PicDRT case study on attitude and perception reporting in pictorial narratives. First, we briefly suggest that the familiar speech and thought bubbles in modern comics may be thought of as the pictorial analogue of quotation, but we leave a detailed study for future research. We then spend some time reconstructing and adapting ideas from Abusch and Rooth (2017) on free perception sequences in PicDRT. Though closely related, our analysis is more pragmatic in nature than theirs. For non-veridical free perception cases, for instance, we invoke a pragmatic mechanism based on Eckardt's (2014) Cautious Update, introducing a pragmatically inferred *de se* perceptual experience operator. We go on to consider a phenomenon we call blended panels, where a protagonist's mental state may be depicted alongside the experiencing protagonist herself in a single panel. We compared this phenomenon to free indirect discourse in linguistic narrative but concluded that the analogy fails. A blended picture is more like a regular indirect attitude report, i.e. involving an intensional operator but no perspective shift or quotation.

In conclusion, applying the formal semantic toolkit of DRT, and importing insights from the semantic analysis of linguistic attitude reporting and quotation, has helped us better understand certain types of attitude representation in the visual domain. At the same time, the discussion of free perception panels in particular has led us to consider the way we represent perception in linguistic narratives, an area that is quite underdeveloped in formal semantics (Hinterwimmer 2017 being a notable exception).

References

Abusch, D. (2012). Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences. *Proceedings of Sinn und Bedeutung 17*.

Abusch, D. (2014). Temporal succession and aspectual type in visual narrative. In L. Crnc and U. Sauerland (Eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pp. 9–29. MITWPL.

Abusch, D. (2015). Possible worlds semantics for pictures. In *The Blackwell Companion to Semantics*.

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

Altshuler, D. and E. Maier (2018). Death on the Freeway: Imaginative resistance as narrator accommodation. In I. Frana, P. Menndez-Benito, and R. Bhatt (Eds.), *Making Worlds Accessible: Festschrift for Angelika Kratzer.* Amherst.

Asher, N. and A. Lascarides (2013). Strategic conversation. *Semantics and Pragmatics 6*(0), 2–1–62.

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

Bray, J. (2007). The dual voice of free indirect discourse: a reading experiment. *Language and Literature 16*(1), 37–52.

Brinton, L. (1980). Represented Perception - a Study in Narrative Style. *Poetics 9*(4), 363–381.

Cohn, N. (2013a). Beyond speech balloons and thought bubbles: The integration of text and image. *Semiotica 2013*(197), 35–63.

Cohn, N. (2013b). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* A&C Black.

Eckardt, R. (2014). *The Semantics of Free Indirect Speech. How Texts Let You Read Minds and Eavesdrop.* Leiden: Brill.

Frege, G. (1892). ber Sinn und Bedeutung. *Zeitschrift fr Philosophie und philosophische Kritik 100*(1), 25–50.

Geurts, B. (1999). *Presuppositions and Pronouns.* Amsterdam: Elsevier.

Giardino, V. and G. Greenberg (2015, March). Introduction: Varieties of Iconicity. *Review of Philosophy and Psychology 6*(1), 1–25.

Greenberg, G. (2013, April). Beyond Resemblance. *Philosophical Review 122*(2), 215–287.

Hinterwimmer, S. (2017). Two kinds of perspective taking in narrative texts. *Semantics and Linguistic Theory 27*(0), 282–301.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal Methods in the Study of Language*, pp. 277–322. Amsterdam: Mathematical Centre Tracts.

Kamp, H. (2016). Articulated Contexts. Unpublished Ms., Stuttgart/Austin.

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind and Language 30*(3), 235–273.

Maier, E. (2017). The pragmatics of attraction: Explaining unquotation in direct and free indirect discourse. In P. Saka and M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation.* Berlin: Springer.

McCloud, S. (1993). *Understanding Comics.* Tundra Publishing.

Saraceni, M. (2003). *The Language of Comics.* Psychology Press.

Schlenker, P. (2004). Context of thought and context of utterance: a note on Free Indirect Discourse and the Historical Present. *Mind and Language 19*(3), 279–304.

Schlenker, P. (2017). Outline of music semantics. *Music Perception: An Interdisciplinary Journal 35*(1), 3–37.

Schlenker, P. (2018). Iconic pragmatics. *Natural Language & Linguistic Theory 36*(3), 877–936.

Zimmermann, T. E. (2016). Painting and opacity. In F. et al. (Ed.), *Von Rang und Namen: Philosophical Essays in Honour of Wolfgang Spohn*, pp. 425–453. Mentis Verlag.