# Temporal and intensional pictorial conflation[1]

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

**Abstract.** This paper proposes accounts in possible worlds semantics of temporally conflated pictorial narratives, and intensionally conflated pictorial narratives. The semantics for film shots and panels in comics that are embedded under attitudes such as imagining and dreaming uses de se interpretation and existential modality, strengthened by a normality condition.

**Keywords:** attitudes, comics, conflated narrative, continuous narrative, de se, discourse representation, film, indexing, modality, possible worlds, premise semantics

## 1. Introduction

*Temporal conflation* as we term it is the phenomenon of a single picture combining information from two or more time points. The painting in (1) is understood narratively, with a bear walking through a ravine up to a tree, where it is shot by a hunter and tumbles down. The bear is shown five times. Masaccio's *Tribute Money* in (2) depicts a New Testament story where a tax collector (in orange) asks a group including Jesus and St. Peter for payment. On the left St. Peter (in blue) takes a coin from the mouth of a fish near a lake, and on the right, St. Peter pays the tax collector. St. Peter is depicted three times, and the tax collector is depicted twice.

(1)

*Maharaja Fateh Singh Hunting Female Bears* (detail).
Attributed to Pannalai.
Metropolitan Museum, New York.

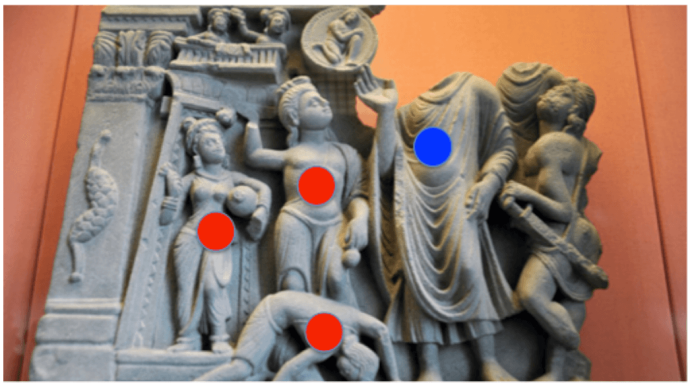
(2)

*The Tribute Money*.
Masaccio.
Brancacci Chapel, Florence.

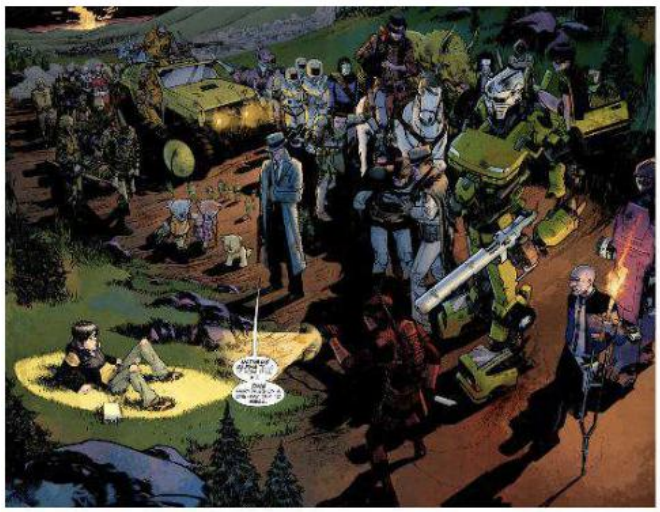
[1]We thank the audience at Sinn und Bedeutung and participants in semantics seminar at Cornell for comments and reactions. Images in the paper that are quoted from comics and film are used for educational and critical purposes, and are property of their respective owners.

Dehejia (1997) discusses examples of temporal conflation in Indian sculpture. The sculpture reproduced in (3) depicts a story where a character Sumedha (marked with red dots) first buys some lotuses. Then he tosses the lotuses at a character Dipankara (an incarnation of the Buddha, marked in blue). Finally Sumedha spreads his hair on the ground for Dipankara to step on. Sumedha is sculpturally depicted three times, just as St. Peter is depicted three times in (2). Dehejia points out the special property that a single depiction of Dipankara seems to figure in two depicted events, the lotus-tossing event and the hair-spreading event, which are understood as temporally sequenced.[2]

(3)



(4)



Bimpikou (2018) and Maier and Bimpikou (2019) discuss examples from comics that potentially are modal or intensional versions of conflation. (4) is from Grant Morrison and Sean Murphy's comic *Joe the Barbarian*. As described in Bimpikou (2018), in this panel "we see a character surrounded by his own hallucinations: here, Joe's toys have come to life. But the character is also depicted, which implicates that a narrator can 'see' both the character and his hallucinations from some external position". So we can say that the panel conflates geometric

[2]Dehejia uses the term "conflated narrative" for visual narratives with this special property. Examples such as (1) and (2) are *continuous narratives* in art historical terminology, e.g. Andrews (1995). This paper uses the term "conflated" for both, to emphasize that the picture or sculpture combines geometric information from two or more world-time pairs.

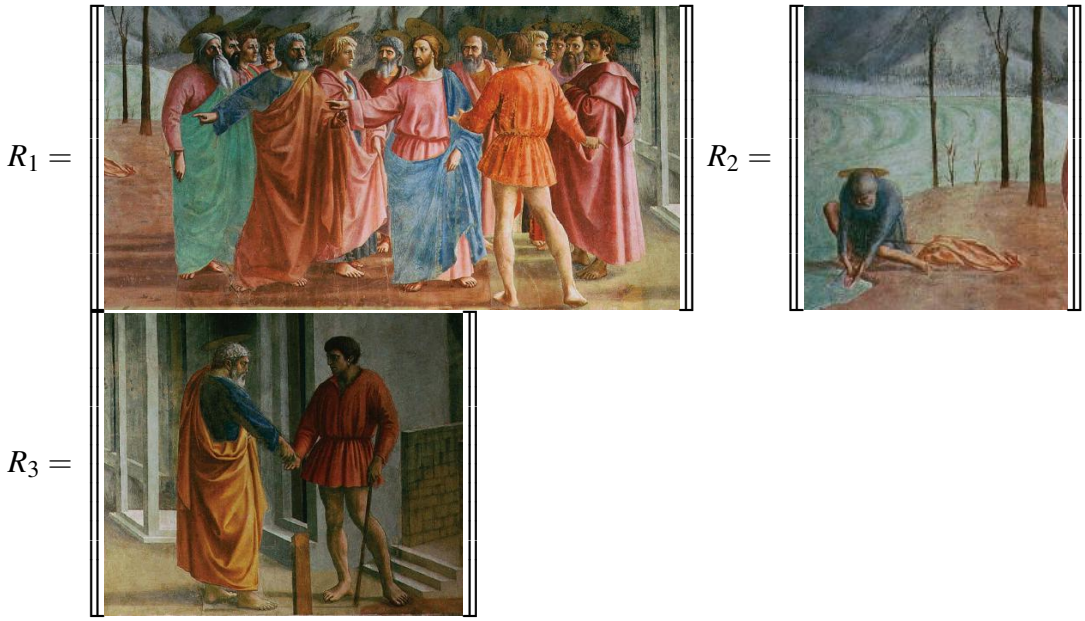



Figure 1: Semantic values from three parts of *The Tribute Money*.

information about Joe in the base world of a described situation, and geometric information about imagine-world alternatives for Joe in the base world of a described situation. In a base world, Joe is in a posture consistent with the bright part of the picture around the depiction of Joe in the panel, he is wearing clothing that looks like that part of the picture from a certain viewpoint, and so forth. In imagination-world alternatives for Joe, there are individuals, animals, and weapons that look like the part of the panel outside the bright part around the depiction of Joe from a certain viewpoint.

This paper develops an approach to temporal conflation that is based on first constructing a three-dimensional space by combining parts of 3D spaces from different times in a world-time line. Then this composite 3D space is projected to a picture (Section 2). This method is then extended to the modal case. We argue that the modal analysis, while attractive for some examples, does not cover the full range of intensional constructions in pictorial narratives (Section 3). And especially for film, there is a systematic problem of the semantic contents of embedded passages being too strong for a standard attitudinal semantics using universal quantification over attitudinal alternatives to work. This problem is not alleviated by spatial conflation. Section 4 develops a solution using weakened quantificational force and a normality condition. In a certain sense, the normality condition takes over the work of combining information from base and attitudinal worlds.

## 2. 2D and 3D splicing

A projective possible worlds semantics for pictures is assumed, following Greenberg (2011). Abusch and Rooth (2017) and Rooth and Abusch (2018) give the following formulation which retains a viewpoint in the semantic value.[3] Let $w$ be a possible world at a time, and let $v$ be a "viewpoint", a package of information that combines a location in space with information that locates an oriented, clipped picture plane. $\pi$ is a projection function that maps a world and a viewpoint to a picture, $\pi(w,v) = q$. Then the semantic value of a picture is defined by inverting projection, $[\![q]\!] = \{\langle w,v\rangle | \pi(w,v) = q\}$. This is a relation that holds between a world and a viewpoint if and only if the world looks like the picture from the viewpoint.

Suppose semantic values for three parts of *The Tribute Money* are defined as in Figure 1. $R_1$, $R_2$ and $R_3$ are each semantic values of parts of the picture.[4] Each part is a relation between worlds and viewpoints. The relations can be pieced together as in (5) to give an interpretation of the picture as a whole, with temporal sequencing. The notation $w < w'$ indicates that $w'$ is a temporal extension of $w$.[5] This is an interpretation that is equivalent to the sequenced narrative (6).

(5)
$$\left\{ \langle w_3, v_3\rangle \middle| \exists w_1 v_1 w_2 v_2 \left[ \begin{array}{l} w_1 < w_2 \wedge w_2 < w_3 \wedge \\ R_1(w_1,v_1) \wedge R_2(w_2,v_2) \wedge R_3(w_3,v_3) \end{array} \right] \right\}$$

(6)

To implement this strategy for temporally conflated narratives, we propose as a starting point the logical form (7), where $q$ is the conflated picture, and $\vec{y}$ is a vector of mathematical information identifying the parts. The order in $\vec{y}$ is significant, because it corresponds to temporal order in worlds of the semantic value. *Cn* is the operator that introduces temporal conflation.

(7) $Cn(q,\vec{y})$

The notion of part needs to be general, because in the example with the bears, a division roughly like in Figure 2 is needed, where the superimposed numbers indicate order in $\vec{y}$. We require that the parts indicated by $\vec{y}$ partition the picture. In *The Tribute Money*, the division is not

[3] Abusch (2021) is a handbook review of the framework.

[4] The panel divisions need to be adjusted so that they partition the painting. See below.

[5] As a working framework we have in mind a branching time model structure, where a "world" has total information about its past, but no deterministic information about its future, and worlds are temporally ordered by a partial order of temporal extension. Model structures combining time and modality can be defined in different ways though (Thomason, 1984).

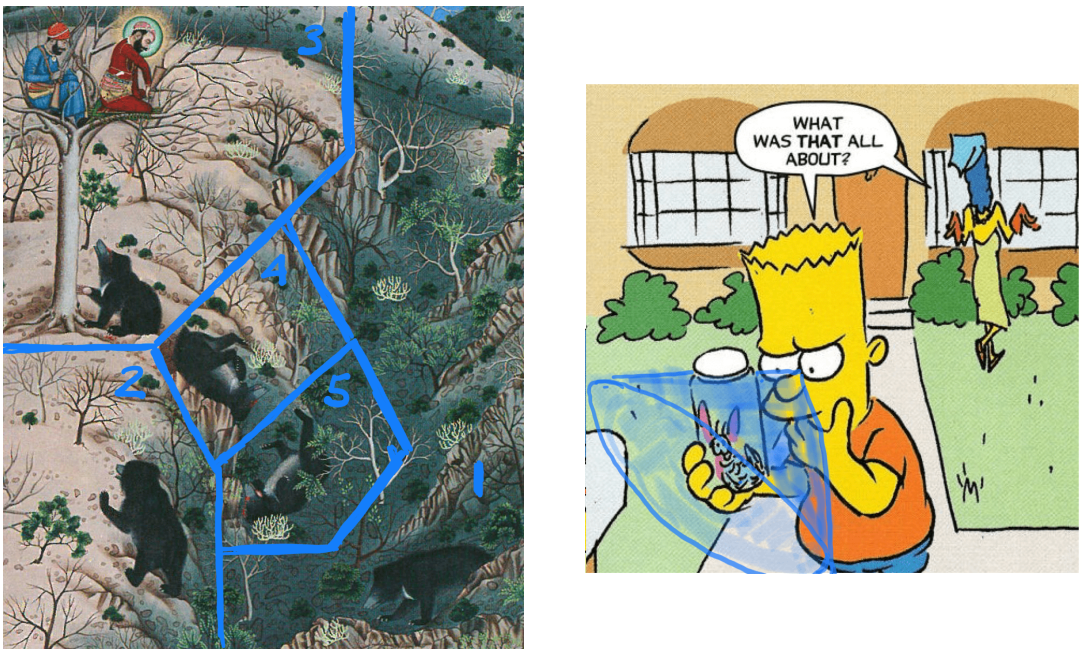



Figure 2: Left: Division of a temporally conflated picture into areas, with ordering corresponding to temporal ordering in the described situatition. Right: Division of a modally conflated volume into two parts, one corresponding to the cone of vision of a character.

trivial, because a division that respects the shadows is required. Notice that the shadow of the depiction of St. Peter in the right part (the third part indicated in $\vec{y}$) extends up to the right leg in the depiction of the tax collector in the center part (the first part). Therefore simple vertical divisions are not adequate, because they cut parts of shadows off from the figures that cast them, and so result in the shadows giving information about the wrong times. There are also issues of illumination in the parts being consistent or not. We would like to be able to say more here, but do not do so, because of the need to take into account more of the corpus of examples in Renaissance European art, Indian art, and elsewhere, and the literature on them.[6]

The LF (7) needs to be incremented with introductions of discourse referents, and equalities between discourse referents that, in the LF of *The Tribute Money*, identifies the three depictions of St. Peter, and the two depictions of the tax collector. This is accomplished by adding geometric points that introduce discourse referents to (7). Let $a$ and $b$ be geometric points that are within the depictions of St. Peter and the tax collector, respectively, in the middle part. Let $a'$ be a geometric point that is within the depiction of St. Peter on the left. Let $a''$ and $b''$ be points that are within the depictions of St. Peter and the tax collector, respectively, on the right. These are gathered together into $[[a,b],[a'],[a'',b'']$, a list of geometric points of the same length as $\vec{y}$. This vector is included as a third argument of *Cn*, and the discourse referents are equated external to *Cn*, in the way shown in (8). Equations between discourse referents use a recency convention (Abusch, 2021). "1=4" identifies the two depictions of the tax collector, because $b''$ is the most recent introduction of a discourse referent, and so the tax collector on the right is referenced with index 1. $b$ is four steps back in introductions of discourse referents,

[6] A bizarre possibility is the shadow of St. Peter on the right falling on the leg of the depiction of the tax collector in the middle. This entails information from two times contributing to a single part of the picture.

and so the tax collector in the center is referenced with the index 4. Identity predications such as "1=4" express identity in the model, and it is assumed that the model has information about identity of individuals across time, expressed as mathematical identity in the underlying set of individuals of the model. All of this leads to the LF (8) for *The Tribute Money*, where $q$ is the basic picture. $y_1$, $y_2$, and $y_3$ are pieces of mathematical information that identify the medial, left, and right parts respectively. In general the LF of conflated pictures is as in (9), where $\vec{z}$ has the same length as $\vec{y}$, where it is required that each point in the list $\vec{z}_i$ is within the area in $q$ identified by $\vec{y}_1$.

(8) $Cn(q,[y_1,y_2,y_3],[[a,b],[a'],[a'',b'']])$ 1=4 2=3 3=5

(9) $Cn(q,\vec{y},\vec{z})$

This scheme treats conflated pictures as sequences of ordinary pictures with a stable viewpoint, with indexing treated as for ordinary pictures in sequential pictorial narratives. The scheme is in many ways attractive. Andrews (1995) points out that viewers of pictures, and in particular large frescoes such as *The Tribute Money*, do not take in the whole thing at once. A viewer who looked at the fresco with glance order following temporal order in the described situation would look at the central part first, then the left part, then the right part. Thus one can say that $\vec{y}$ in the LF indicates intended glance order. Sometimes depictions of pointing, like Jesus and St. Peter pointing towards the lake, give clues about $\vec{y}$.

In an alternative approach to temporally conflated narratives, information from different times is combined three-dimensionally, rather than two-dimensionally. Consider the spatial volume $Y$ in a described situation for the fresco. This space can be divided into a volume $Y_2$ by the lake where St. Peter takes the coin from the fish mouth in the second episode, the spatial volume $Y_3$ near the building where St. Peter pays the tax collector in the thirde episode, and the spatial volume $Y_1$ where in the first episode Jesus and his disciples interact with the tax collector. We assume that the volumes are defined relative to a viewpoint and assume coordinates relative to the earth that make the viewpoint stable across time.

$Y$ is the combination of the three spatial parts, $Y = Y_1 \cup Y_2 \cup Y_3$. We outline a method for projecting a world $w$, given a viewpoint $c$ and vector of such three-dimensional volumes (located relative to a viewpoint) which we exemplify with $[Y_1,Y_2,Y_3]$. Non-deterministically pick initial segments $w_1$ and $w_2$ of $w$, $w_1 < w_2$ and $w_2 < w$. Where $Y'$ is a spatial volume, $w'$ is a world, and $v'$ is a viewpoint, let $G(Y',w',v')$ be the geometric information in $w'$ about what individuals are present in $w'$ within the volume $Y'(v')$, the configuration and orientation of those individuals, and so forth. A geometric data structure or "world" that combines parts from $w_1$, $w_2$, and $w$ is then formed as $G(Y_1,w_1,v') + G(Y_2,w_2,v') + G(Y_3,w,v')$, where the sum combines geometric information from the spatial parts. Then this composite world is projected to a picture in the normal way.[7] In summary then, a geometric world is formed by combining different spatio-temporal parts of a world, and that composite world is projected to a picture.

[7]While the description here is high-level, all of this can be realized concretely using the data structures used for rendering films from data structural descriptions of world-time lines in systems such as Blender (Blender Foundation, 1994). In this setting, $G(Y_i,w_i,v')$ is a list data structure listing individuals, their configurations (e.g. the angle of joints in the case of people), locations, and orientations.

The spatial scheme for temporal conflation is potentially more flexible, and even simpler, in some cases. Consider a film shot in which a character looks out through a window, and sees her younger self on the street.[8] Sometimes the depiction of the street scene and the younger character is occluded by the depiction of the older character. In the method using spatial divisions, one can assume a volume $Y_1$ outside the house, and a volume $Y_2$ inside the house, and use the same spatial division $[Y_1, Y_2]$ to project each film frame. An analysis using 2D-divisions cannot use stable divisions, because the scene at earlier time is partially occluded by the depiction of the older character, and that character moves.

The LF for projection using spatial divisions is nearly the same as the one for 2D-divisions. In (10), $\vec{Y}$ is a vector of mathematical objects that identify spatial volumes relative to a viewpoint. The volumes are assumed to partition space. The geometric points $\vec{z}$ are defined relative to the 2D picture, as before.

(10) $Cn(q, \vec{Y}, \vec{z})$

This completes our presentation of LFs for temporal conflation. We presented one version which uses 2D divisions, and another with 3D divisions. In each case, the LF includes information that locates the divisions, and their temporal order. We outlined interpretations of the LFs that deterimine whether a given world and witness sequence satisfy the conflated picture.

## 3. Modal conflation

This section considers the possibility that pictures in pictorial narratives and shots in film can be *modally* or *intensionally* conflated. Instead of information from multiple times in one world, the picture combines information from different worlds which are not part of the same world-time line. Abusch and Rooth (2017), Bimpikou (2018), Rooth and Abusch (2018), Maier and Bimpikou (2019), and Abusch and Rooth (2021) have analyzed attitudinal constructions in pictorial narratives in intensional semantics. One case of this is "free perception" or "point of view" constructions that express the perceptual experiences of characters, or the epistemic states of characters that result from experiences of looking. (11) is a free perception sequence from the short comic story *Blood Curse of the Evil Fairies*, with characters from the Simpsons (Baker, 2006). The middle setup panel shows Bart looking at an empty jar, with his sister Lisa in the foreground. The third panel shows a creature with wings, a fairy. It is understood that in a described situation for the story, Bart sees or hallucinates a view like the third panel, and/or is in an epistemic state that results from seeing or hallucinating such a view. The first panel makes clear that Lisa does not see the fairy.

[8] This is inspired by the initial passage of *Antonia's Line*, which however uses a point of view shot (Gorris, 1995). In the memory shot, the older incarnation of Antonia is not visible. We conjecture though that shot described in the text is possible.

(11)


The second and third panels in (11) are a free perception or point of view sequence, with the one panel showing an agent, and the other panel showing what they see or hallucinate. Abusch and Rooth (2017) and Abusch and Rooth (2021) analyzed free perception sequences in an epistemic possible worlds and event semantics framework, hypothesizing that an intensional reading of free perception had an embedding operator in the LF or discourse representation. The semantics in those papers refers to how worlds appear from the perspective of agents in alternative worlds that are compatible with the attitudinal state of the agents. The LF or discourse representation for a free-perception sequence $p\ q$ with an intensional interpretation is as in (12). The setup picture $p$ is unembedded, and has an extensional interpretation. $x$ introduces a discourse referent for the agent depicted in $p$. The second picture $q$ is embedded under an attitude predicate $A$, the agent of which is the discourse referent introduced by $x$.[9] In the semantics, the counterpart of the agent in a centered world $\langle w', v' \rangle$ satisfying $q$ is an agent with the geometric viewpoint $v'$.

(12) $p\ x\ A_1(q)$

We think it is correct to analyze free perception as a distinct construction with a syntax and semantics that is specific to that construction. However, Bimpikou (2018) pointed out that there are panels in comics that reflect what a character hallucinates or imagines, but which are projected from a neutral viewpoint, rather than the viewpoint of the character whose attitudinal state is being described. The panel (4) from *Joe the Barbarian* is an example. The second panel in the sequence (13) from *Blood Curse of the Evil Fairies* is another. This panel is discussed in the same theoretical context in Maier and Bimpikou (2019). It shows Bart looking at a jar containing a fairy, with the interpretation that Bart sees the fairy. Bart's mother Marge is in background. She has turned away from Bart and the jar.

[9] In the formulation from Abusch (2012), $x$ is an area, the projection of the agent in picture $p$. In the formulation from Abusch (2021), $x$ is a point within the projection of the agent. Either way, $x$ introduces a discourse referent with a geometric constraint on it, and that discourse referent is referenced with the index 1, using a recency convention as in Dekker (1994).

(13)


Bimpikou (2018) outlined an approach to neutral-viewpoint attitudinal panels involving two-dimensional splitting. It is similar to the approach using 2D splitting of temporally conflated narratives from Section 2. The panel is split into a part that reflects the information of the agent, and another part that gives information about the base world. Bimpikou points out that this makes a modally conflated panel parallel to a panel with a pictorial thought bubble, both in its syntax and in its semantics. (14) is her version of the example from *Joe the Barbarian* with a pictorial thought bubble, interpreted as showing what the character Joe imagines.

(14)

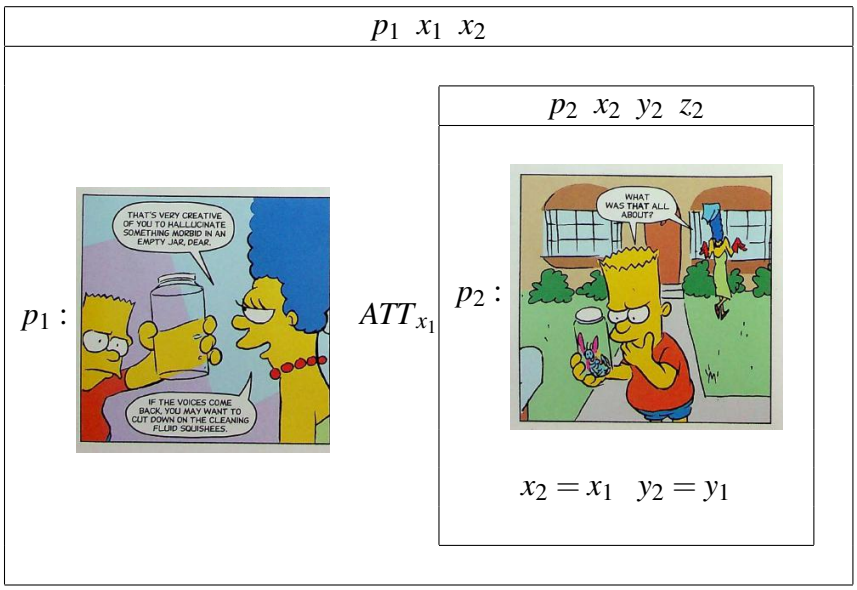
From Bimpikou (2018).

As an alternative to 2D splitting, consider an analysis where splitting is accomplished in three-dimensional space. For (13), we have in mind that the second picture is projected from a three-dimensional space that is separated into a solid cone corresponding to Bart's visual pyramid, which has information about Bart's attitudinal state, and a complement 3D volume that is a sub-part of the base world. See the image on the right in Figure 2.

Maier and Bimpikou (2019) propose a substantially different analysis of the fairy panel, where all of the panel is embedded, and is interpreted as giving information about the attitudinal state of the agent. They propose the representation (15), given in the box syntax of DRT.[10] There is an embedding predicate *ATT* that embeds all of the picture with the fairy in the jar, and there is no splitting, either in two or three dimensions.

[10]The DRT notation (15) can be transposed to a linear notation like (12), also with an embedding predicate.

(15)




The semantics for the DRS (15) that Maier and Bimpikou suggest is a universal quantification over attitudinal alternatives. Any attitudinal alternative world for the agent $x_1$ in a base world that verifies the DRS as a whole is required to satisfy the embedded DRS that contains the fairy picture.[11] This is potentially subject to two objections. First, one might want to claim that the part of the picture containing the mother, showing her with her back turned to Bart, and with her body and limbs in such-and-such orientation, gives information about the mother in the base world. Relatedly, the embedded picture seems to have too much information for a universal quantification over alternatives for Bart to be satisfied in a plausible model. If Bart is facing away from his mother, he cannot see his mother, and does not have perceptual information about the configuration of her body and limbs. Presumably, in a world consistent with the picture, Bart has world-alternatives where Marge's legs are in a variety of different postures, including crossed (as in the picture), and non-crossed. If so, the condition that all world alternatives for Bart look like the fairy picture (from some viewpoint) is not satisfied, because Bart has world alternatives where the mother's legs are not crossed, while the mother's legs are crossed in all worlds consistent with the picture.

The phenomenon of intensional passages in graphic narratives assuming a neutral viewpoint, rather than the geometric viewpoint of a character, is pervasive. Here are two examples from film. In Episode 8, Season 3 of *Poldark*, the protagonist meets an old flame Elizabeth in a churchyard, talks to her, and kisses her. In the subsequent passage, he is shown talking to his wife Demelza, and confessing the churchyard incident to her. The confesson has an interpretation of imagination. The confession shot shows Poldark talking, with a neutral viewpoint. See (16). In the film *Gravity*, a scientist-astronaut Stone finds herself in orbit in a Soyuz capsule, with low supplies of oxygen and fuel, and apparently no means of returning to earth. She decides to end things, turns off the oxygen, and closes her eyes. Thereafter, in what is interpreted as a dream or reverie of Stone's, the seasoned astronaut Kowalski (who had apparently died during a spacewalk earlier in the film) enters the capsule from outside, talks to Stone in a way that

[11] It can be stated like this: $[\![ATT_{x_1} K]\!]^{f,v,w} = 1$ iff for all $w'$ compatible with $f(x_1)$'s attitudinal state in $w$ : $[\![K]\!]^{f,v,w'} = 1$.

raises her spirits, and suggests a way of using fuel in an auxiliary module to brake the capsule and return to the surface. While the passage has the interpretation of a dream or hallucination of Stone's, the camera angle for the passage is not identified with the geometric perspective of the character Stone. Rather the passage is shot from a neutral perspective, showing Stone and Kowalski and the interior of the capsule. See (17).

(16) *Poldark*, Episode 3, Season 8 (Barry, 1975).
Poldark: I met Elizabeth. For the first time in years, we talked.

(17) From *Gravity*, Cuarón (2013) Scene 7



These intensional filmic passages raise the same problem as the fairy panel, but in more extreme form. The semantic content of the film shots is very strong. The passage from *Gravity* is three and a half minutes long, and at each time in this interval, Stone's body is shown in a particular posture, and her face has a particular projected appearance. Similarly for Kowalski. Kowalski's and Stone's utterances have particular detailed acoustic waveforms. In the background, there are straps and folded fabric in a particular configuration. There is writing on the instrumentation in the Cyrillic alphabet. There are dynamic displays with flashing lights, which in the film shot have a particular time course. An attitudinal semantics using universal quantification has it that Stone's attitudinal state in the base world for the film (her set of dream-alternatives or reverie-alternatives) has a content that entails the strong content of the shot of Kowalski entering the capsule. It is completely implausible that the film describes Stone as dreaming something this specific.

Thus, the results of our experiment support the conclusion that there is a genuine difference between conjunction and disjunction. Above we said that a semantics using modal conflation using spatial divisions was potentially a good analysis of the fairy panel, in that it allowed the appearance of the jar in the picture to constrain Bart's attitudinal state, and the appearance of the mother in the picture to constrain the base world. Observe that this kind of analysis is in no way attractive for the *Poldark* and *Gravity* passages. Where Kowalski is partially out of view of Stone in part of the passage, it does not follow that Kowalski is present in the capsule in the base world. In the *Poldark* passage, Poldark is not in his own view, and it is implausible to maintain that Poldark is being described as assuming in the base world the facial posture that is shown in the shot. We conclude that an approach involving two-dimensional or three-dimensional splitting does not cover the full range of examples, and that a more general solution not involving geometric conflation is needed.

In Abusch and Rooth (2021), considerations of embedded pictures in point-of-view constructions having over-strong contents led us to make the move of replacing universal quantification in the attitudinal semantics with existential quantification. Without qualification, this is unacceptably weak, and we added additional constraint having to do with normal perceptual interactions of agents with their environments. Roughly, a free perception sequence with a setup picture $p$ and an embedded picture $q$ is satisfied if and only if the agent $x$ depicted in $p$ takes a perceptual action $e$ (which is an event of hallucinating), some alternative to which is a perceptual action $e'$ that the counterpart of the agent could normally take while facing a scene like $q$. We argued that this weakened the information attributed to the agent in the base world, without trivializing it. The next section takes a similar tack to intensional passages with a neutral viewpoint.

## 4. An existential modal semantics for embedding

We have in mind this kind of model structure for events of dreaming and imagining. When Poldark imagines the confession, or when Stone dreams the scene with Kowalski, they participate in an imagination or dreaming event $e$ in the base world of the narrative. These events encode agents, so that given an event one can recover the agent of the event. Events of dreaming and imagining have contents, which are agent-centered propositions, with the content of event $e$ written $C(e)$. In an element $\langle w', x' \rangle$ of $C(e)$, $x'$ is conceived of as a counterpart of the agent of $e$, the "self", and $w'$ is a world where what the agent imagines has just happened. For instance, if $e$ is an event of imagining seeing a dragon and $\langle w', x' \rangle \in C(e)$, then in world $w'$, agent $x'$ has just seen a dragon. $C(e)$ is analogous to a set of centered epistemic alternatives, but plays a different role, because imagining seeing a dragon is different from believing one is seeing a dragon.

In this discussion, events are token events that occur at just one world/time, so it is not necessary to refer to worlds in specifying properties of events such as what the content is and who the agent is. To distinguish dreaming from imagining and other classes of attitudinal events, the model struture specifies disjoint sets of dreaming events, imagining events, and so forth.

Attitudinal film shots like the ones in (16) and (17) are different from point of view shots in a free-perception construction, in that the viewpoint or camera position is not identified with

the geometric viewpoint of the de se character. Rather, the viewpoint is neutral. It does not follow that the information conveyed by neutral-perspective panels/shots is fundamentally different from the information conveyed by point-of-view panels/shots. In models with attitudinal events, both kinds of embedded passages provide information about the content of an attitudinal event that happens in the base world. In a point of view shot, the self is an agent whose visual-geometric position coincides with the viewpoint for the embedded shot. To achieve similar ends in neutral-viewpoint passages, we assume that the embedded passage contains a distinguished discourse referent for the self, which is present syntactically in the embedding construction. By convention, this is the last discoure referent introduced. For instance, where *Dr* is the dreaming predicate and $q$ is the shot in *Gravity*, the formula $Dr_i(q\ a)$ embeds $q$ under an dreaming predicate with agent (or subject) $i$, and with $a$ introducing a discourse referent for the counterpart of Stone in the embedded context.

Explanation is required about the nature of geometric discourse referents when $q$ is a film shot. Abusch (2021) treated the case of pictorial narratives such as comics, and there a discourse referent is introduced by a point in the picture.[12] This does not work for film shots, which consist of a sequence of frames, rather than a single panel. For this case, the geometric discourse referent $a$ is a *partial sequence* of locations, indicating the location of the self throughout the shot. The sequence is partial, because in some frames of the embedded shot, the self may be out of view. Using a superscript to mark frames of the shot, a witness for $a$ in the configuration $q\,a$ is an individual whose projection in frame $q^j$ of $q$ surrounds point $a^j$, at each index $j$ such that $a^j$ is defined, i.e. in each frame where the self is in view.

A starting point for discussion of the semantics of embedding in pictorial narratives is the analysis in Lewis (1979) of de se attitudes. As applied to *believe*, this can be described as hypothesizing that (i) the belief state of an agent in a world is characterized by an agent-centered proposition; (ii) by some systematic means, an agent-centered proposition is obtained from the complement of *believe*; and (iii) a sentence "$x$ believes that $\varphi$" is true in world $w$ iff the agent-centered proposition that characterizes the beliefs of $x$ in $w$ entails the agent-centered proposition obtained from $\varphi$. We already said that an imagination or dreaming event has an agent-centered proposition as a content; this will be used in place of (i). For (ii), it is necessary to extract an agent-centered proposition from an embedded pictorial narrative. Given the assumptions about de se discourse referents, this is accomplished by (18), which collects pairs $\langle w', x'\rangle$ such that $w'$ together with a witness sequence $\mathscr{O}'$ for discourse referents and some viewpoint satisfy the embedded narrative. The condition $x' = \mathscr{O}[1]$ uses the convention that the de se discourse referent is introduced last, so that the self is referenced with index 1.

(18) The agent-centered proposition $\bar{C}(\varphi)$, where $\varphi$ is of the form $qa$, is the set of pairs $\langle w', x'\rangle$ such that for some witness sequence $\mathscr{O}'$ and viewpoint $v'$, $\langle w', v', \mathscr{O}'\rangle$ satisfies $\varphi$, and $x' = \mathscr{O}'[1]$.

It is now a simple matter to state an entailment semantics for attitudinal embedding in pictorial narratives. (19) requires that for $Dr_i(\varphi)$ to be true, the content of a dreaming event in the base world with agent $i$ must entail the content extracted from the complement. Both contents are agent-centered propositions, and entailment for agent-centered propositions is subset.

[12] Or in variant formulations, an area of the picture or a bounding box.

(19) *Entailment semantics for dreaming in pictorial narratives*
$w, \mathcal{O}$ satisfies $Dr_i(\varphi)$ iff there is a dreaming event $e$ that has just occurred in $w$, $e$ has agent $\mathcal{O}[i]$, and $C(e)$ entails $\bar{C}(\varphi)$.

This semantics is certainly too hard to satisfy for embedded examples in film. The reason was stated above—the information in film shots is geometrically and acoustically very specific, and it is completely implausible that the information of base-world events such as Stone's dreaming event should be specific enough for its content to entail this very strong film-shot content.

This point is clear for acts of imagining which are not visual imagining. Consider a scenario where in the base world, Poldark simply repeats the confession silently, word by word, but without imagining anything visual or geometric. The information content of such an event sequence is certainly not strong enough to entail the visual-geometric content of the *Poldark* passage. It is also not strong enough to entail the phonetic and acoustic detail in audio for the shot. Yet we think that the confession shot is consistent with this kind of low-information, non-visual imagining.

Thus there is a fundamental mismatch between the relatively weak informational content of events of imagining or dreaming, the strong informational content of film shots and realistic pictures, and a semantics for attitudinal constructions requiring entailment. We propose addressing the incompatibility in a radical way, by replacing entailment with compatibility. If it is merely required that what is shown in the embedded shot is *compatible* with what Poldark imagined, we do not get the consequence that Poldark imagined the acoustic waveforms of the words he imagined speaking.

We give a semantics for attitudinal embedding in pictorial narratives in existential form, in that for an imaginaton formula $Im_i(\varphi)$ to be true, there must be a witness centered world $\langle w', x' \rangle$ that satisfies certain conditions. One of the conditions is roughly that $w'$ looks like the embedded shot in the narrative. Another is that $w'$ is consistent with the content of an imagining event that, in the base world, the agent picked out by $i$ has just participated in. In the definition (20), part (iii) says that the centered world $\langle w', x' \rangle$ is in the content of the event $e$ in the base world. Part (iv) says that it is in the centered proposition extracted from the complement. The definition for a dreaming predicate *Dr* is the same, except that in (i), $e$ is required to be a dreaming event rather than an imagining event. It is assumed that the model specifies disjoint sets of imagining events, dreaming events, and so forth.

(20) Existential semantics of *Im*, first version.
$Im_i(\varphi)$ is true relative to world $w$ and witness sequence $\mathcal{O}$ iff there is an event $e$, a world $w'$, an individual $x'$, and a witness sequence $\mathcal{O}'$ such that the following conditions are satisfied.
(i) $e$ is an imagining event that has just happened in $w$.
(ii) The agent of $e$ is $\mathcal{O}[i]$.
(iii) $\langle w', x' \rangle$ is an element of $C(e)$.
(iv) $\langle w', x' \rangle$ is an element of $\bar{C}(\varphi)$.

The move to an existential semantics solves the problem of the embedded passage having strong geometric and acoustic content, because the content of the imagination event is not required to

entail that content, just to be compatible with it. But in this form, the semantics is unacceptably weak. Consider a version of the confession shot which, together with Poldark confessing, has a monster entering the scene from behind and approaching Poldark in a threatening way. There are centered worlds $\langle w'', x'' \rangle$ which satisfy this shot, with $x''$ confessing and a monster approaching $x''$ from behind, and which are arguably also consistent with a low-information event of imagining the confession, because in $w''$, the self $x''$ does make confession.

Along somewhat similar lines, the detail about the interior or the capsule in the *Gravity* shot is so rich that Stone and her counterpart can be assumed not to be aware of it. It follows that not all of this detail is entailed by the content of the imagining event in the base world—Stone has not specifically imagined the details. These details are consistent with earlier shots of the interior of the capsule that have an extensional interpretation. This is reminiscent of the spatial splitting approach to the Bart/fairy example, where the portrayed world was split spatially into a part that gave information about the agent's attitude, and a part that gave information about the base world. But for the film shot in *Gravity*, it does not seem right to claim that events that are out of view of the counterpart of Stone—such as particular flashes of lights on instrumentation—necessarily happen in the base world.

A general approach to strengthening the existential semantics in (20) is to require that the witness world $w'$ be normal in certain respects. Certainly, a monster entering Poldark's farmhouse and approaching him from behind is not a normal scenario. And given the situation in the capsule when Stone fell asleep or fell into a reverie, it is normal for the capsule to be generally similar to what it was before. Note though that Kowalski entering the capsule after previously tumbling off into space, in a distant location and without a supply of oxygen, is not normal. What should distinguish it is that Kowalski entering the capsule is part of what Stone dreamt in the base world.

A definition of normality is given in Kratzer's modal framework of premise semantics (Kratzer, 1978, 1981). This definition is applied by Kratzer in ways that strengthens the content of existential modal statements, such as those expressed by *could* in English. Consider Justin and Keisha in their back yard with their dog Snowy. On one occasion Justin playfully tosses a pingpong ball on Snowy's back. Keisha cautions him as in (21a). She uses an existential modal, which in a naive modal semantics expresses existential world quantification. Justin can reasonably object that while anything is possible and there are worlds that are in principle possible where he tosses the pingpong ball and Snowy is seriously hurt, a dog being seriously hurt by a lightly tossed pingpong ball does not happen in any normal course of events. Arguably what Keisha said is false on the relevant understanding. On another occasion, Justin idly starts to stand a big log up vertically in Snowy's vicinity. Keisha cautions as in (21b). Justin could not object in the same way, and what Keisha said is true on reasonable assumptions.

(21)a. Don't do it again. If you do it again, Snowy could be seriously hurt.
b. Don't do that. If the log falls over, Snowy could be seriously hurt.

Kratzer's semantics for such examples refers to two parameters, a "modal base" that is used to characterize the situation in the yard, and an "ordering source" which in this case is something like a rule of thumb theory of backyard physics and dog physiology. Both are sets of propositions. On a simple version of the semantics, a might-conditional is true relative to a modal base

and ordering source if and only if there is some witness world $w'$ where the prejacent is true, the if-clause is true, each proposition in the modal base is true, and there is no competitor world $w''$ where each proposition in the modal base is true, the if-clause is true, and which satisfies a strict superset of the propositions in the ordering source that are satisfied by $w'$. The latter is the characterization of the possible world $w'$ where the ball is thrown and Snowy is seriously hurt being normal.[13]

Including normality radically strengthens the semantics of conditionals with existential modals, meaning it makes them harder to satisfy, and makes them more informative. For (21a) to be true, it is not sufficient that there is some crazy world where Justin tosses the pingpong ball and Snowy is seriously hurt. As a result of increased informativity, in saying (21b) Keisha is saying something non-trivial and informative.

Above, we saw that the semantics (20) using existential world quantification was too weak. We strengthen it in a way that closely parallels the semantics for existential modals in Kratzer's semantics. (22a) is a paraphrase in words of entailment semantics for embedding under *Dr* in the *Gravity* example. The modal is universal, "no matter what" indicates quantification unrestricted by normality, and "exactly" emphasizes that all of the detail in the shot is required to obtain in any world where what Stone dreamt has transpired. Intuitively we think the strong and hard-to-satify truth conditions of the entailment semantics are reflected in the paraphrase (22a). We replace it with (22b), where the modal is existential, and "in a normal course of events" indicates quantification restricted by normality. (23) is a modal paraphrase of an existential semantics for the Poldark example, strengthened by normality.

(22)a. If what Stone dreamt transpired, things would no matter what look and sound exactly like the embedded film shot.

b. If what Stone dreamt transpired, things could in a normal course of events look and sound exactly like the embedded film shot.

(23) If what Poldark imagined transpired, things could in a normal course of events look and sound exactly like the embedded film shot.

Consider the variant of the Poldark shot with the monster entering the farmhouse. Keeping what Poldark imagines in the base world constant, there is arguably no centered world that can serve as a witness for the condition expressed in (23). The reason is that any world $\langle w', x' \rangle$ that satisfies the monster-shot could be made more normal by removing the monster, by standards of normality that capture what tends to happen in eighteenth-century farmhouses. And in $\langle w', x' \rangle$ modified to remove the monster, the embedded shot with a monster is not satisfied. Consider a variant of the *Gravity* shot where the interior of the Soyuz capsule looks different in arbitrary ways from what it looked like in the earlier extensional shots. This might be compatible with Stone's information and the information in her dreaming event. But if the modal base encodes information about the earlier configuration of the capsule, the ordering source captures the

[13] To expand the example, it is in principle possible for a pingpong ball to be accelerated to a dangerously high speed by the random impact of air molecules on one side. And it is in principle possible for even a light impact on a dog's spine to cause a seizure. Worlds where these unusual things happen are assumed not to be normal according to the relevant ordering source. They are not witnesses for the truth of (21a), because there are more normal worlds where the ball is thrown and Snowy is not seriously hurt.

fact that the configuration of the interior of space capsules does not spontaneously change, and given that Kowalski's entering the capsule would not cause such changes, a world with arbitrary changes is not normal, and cannot be a witness to the truth of (22b). Finally, consider the potential problem of worlds where Kowalski enters the capsule being radically abnormal. This is not an issue, because the content $C(e)$ of the base world event contributes the if-clause in the paraphrase (22b). In Kratzer's semantics, the if-clause is in effect added to the modal base, so that worlds that are not consistent with the if-clause are not competitors to witnesses for the existential modal.

Let $\mathscr{N}(w',X,Y)$ notate world $w'$ being normal relative to a set of propositions $X$ serving as the modal base, and a set of propositions $Y$ serving as the ordering source. The modal base and ordering source parameters are in fact functions in Kratzer's framework, which need to be applied to a world to obtain a set of propositions.[14] Where $M$ and $O$ are modal base and ordering source functions, $M(w)$ and $O(w)$ are the corresponding sets of propositions, obtained by evaluating the functions in a base world $w$. Then $\mathscr{N}(w',M(w),O(w))$ expresses that world $w'$ is normal according to modal base functions $M$ and $O$, as evaluated in a base world $w$. $\mathscr{N}(w',M(w)\cup\{p\},O(w))$ expresses that world $w'$ is normal according to modal base functions $M$ and $O$, as evaluated in a base world $w$, with a proposition $p$ (conceived of as coming from an if-clause) added to the modal base.

We would like to strengthen the existential semantics (20) for attitudinal pictorial embedding by adding the normality condition $\mathscr{N}(w',M(w)\cup\{C(e)\},O(w))$ as a conjunct. The content $C(e)$, glossed as "what the agent imagined" is added in the position of the if-clause, to obtain a condition amounting to (22b) or (23). But there is a type mismatch. In premise semantics, the modal base is a set of propositions, while the content $C(e)$ is a centered proposition. This could potentially be fixed by existentially quantifying the selves to obtain the proposition $\{w'|\exists x'.\langle w',x'\rangle\in C(e)\}$ Or it could be fixed by replacing worlds with centered worlds in premise semantics throughout, so that a modal base is a set of centered propositions.

An example shows that the latter course is correct. Consider a complex world $w''$ which in one location has a Poldark-like agent $x_1$ in a farmhouse in a Cornwall-like area, confessing to his wife with a monster approaching from behind, and in another location (in a galaxy and on a planet remote from the first) has a Poldark-like agent $x_2$ in a farmhouse in a Cornwall-like area, confessing to his wife with no monster. We would like to say that a shot of $x_1$ in $w''$ has abnormal information, but a shot of $x_2$ in $w''$ has normal information. But $w''$ is not normal or abnormal in an absolute way.

Accordingly, for application here, premise semantics is revised by replacing worlds with agent-centered worlds throughout. This is straightforward, since it simply involves re-building definitions of propositions, modal bases, ordering sources, and normality with centered worlds rather than worlds at the base. The normality condition now takes the form $\mathscr{N}(\langle w',x'\rangle,M(\langle w,x\rangle)\cup\{C(e)\},O(\langle w,x\rangle))$, where worlds are replaced with centered worlds also in several other places. $x$ is the agent of the base world event. In the revised semantics for *Im* in (24), this condition is added as a constraint on the witness centered world $\langle w',x'\rangle$.

[14] This becomes relevant when modals are embedded, and when one considers the information that an unembedded modal adds to the common ground

(24) Existential semantics of *Im*, second version with normality.
*Im*$_i(\varphi)$ is true relative to world $w$ and witness sequence $\mathscr{O}$ iff there is an event $e$, an individual $x$, a world $w'$, individual $x'$, viewpoint $v'$, and witness sequence $\mathscr{O}'$ such that the following conditions are satisfied.
(i) $e$ is an imagining event that has just happened in $w$.
(ii) $x = \mathscr{O}[i]$
(ii) The agent of $e$ is $x$.
(iii) $\langle w', x' \rangle$ is an element of $C(e)$.
(iv) $\langle w', x' \rangle$ is an element of $\bar{C}(\varphi)$.
(v) $\mathscr{N}(\langle w', x' \rangle, M(\langle w, x \rangle) \cup \{C(e)\}, O(\langle w, x \rangle))$

As before, the semantics of the dreaming predicate *Dr* is the same but with "$e$ is a dreaming event" substituted in (i).

Section 2 proposed an analysis of temporally conflated narrative using spatial splitting, and Section 3 extended it to intensionally conflated examples, referring to the fairy jar example. This section has presented an account of neutral-perspective embedded shots using existential force, normality, and de se interpretation. This is in competition with the spatial-splitting approach from Section 3. We argued that the approach using existential force and normality was more general, since spatial splitting is not an option for examples such as the dream shot from *Gravity*. There is a potential case which would favor spatial splitting. The LF in this section referred to an agent of the embedding relation, and counterparts of that agent introduced by a discourse referent. This entails that there is just one hallucinating agent. Consider a version of the jar panel in which there are two agents, each shown hallucinating different things, each of which is shown in the visual pyramid of the agent. For instance, Bart might be shown hallucinating a fairy in a jar in his hands, and Homer be shown hallucinating a monster in a jar in his hands in another part of the picture.[15] This could be accommodated in the spatial splitting approach, by introducing a vector of spatial divisions, and a vector of discourse referents for hallucinating depicted agents. It is not as easy to accommodate it in the approach from this section. We don't have an example of this character, but are inclined to think it is possible.

## References

Abusch, D. (2012). Applying discourse semantics and pragmatics to co-reference in picture sequences. In E. Chemla, V. Homer, and G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung 17*, pp. 9–25.

Abusch, D. (2021). Possible-worlds semantics for pictures. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. E. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics*, pp. 2329–2359. Blackwell: Wiley.

Abusch, D. and M. Rooth (2017). The formal semantics of free perception in pictorial nar-

[15] In fact in *Blood Curse of the Evil Fairies*, Homer can see the fairy too. In one panel the fairy explains "Incredible! I can't be seen by human adults! Only the simple, innocent mind of a child can perceive ...". There is also a frame story, with Bart reading to his other sister Maggie. These points could be held to indicate that the fairy panel takes the stance of realism about fairies, rather than intensionalism.

ratives. In A. Cremers, T. van Gessel, and F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium*, pp. 85–95.

Abusch, D. and M. Rooth (2021). Modalized normality in pictorial narratives. In P. Grosz, L. Martí, H. Pearson, Y. Sudo, and S. Zobel (Eds.), *Proceedings of Sinn und Bedeutung*, Volume 25, pp. 1–18.

Andrews, L. (1995). *Story and Space in Renaissance Art: the Rebirth of Continuous Narrative*. Cambridge University Press Cambridge.

Baker, K. (2006). Blood curse of the evil fairies. In *Bart Simpson's Treehouse of Horror, No. 12*, pp. 22–33. Bongo.

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

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

Blender Foundation (1994). *Blender*. 1994-2022.

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

Dehejia, V. (1997). *Discourse in Early Buddhist art: Visual Narratives of India*. Coronet Books Inc.

Dekker, P. (1994). Predicate logic with anaphora. In M. Harvey and L. Santelmann (Eds.), *Proceedings from SALT 4*, pp. 79–95.

Gorris, M. (1995). *Antonia's Line*. Asmik Ace Entertainment.

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

Kratzer, A. (1981). The notional category of modality. In H. J. E. und Hannes Rieser (Ed.), *Words, Worlds, and Contexts. New Approaches in Word Semantics*, pp. 38–74. Berlin: de Gruyter.

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

Maier, E. and S. Bimpikou (2019). Shifting perspectives in pictorial narratives. In M. Espinal, E. Castroviejo, M. Leonetti, L. McNally, and C. Real-Puigdollers (Eds.), *Proceedings of Sinn und Bedeutung 23*, pp. 1–15.

Rooth, M. and D. Abusch (2018). Picture descriptions and centered content. In R. Truswell, C. Cummins, C. Heycock, B. Rabern, and H. Rohde (Eds.), *Proceedings of Sinn und Bedeutung 21*, Volume 2, pp. 1051–1064.

Thomason, R. H. (1984). Combinations of tense and modality. In D. Gabbay and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, pp. 135–165. Springer.