Possible Worlds Semantics for Pictures

Dorit Abusch

April 28, 2019

1. Introduction

In the philosophy of language and linguistic semantics, there is an approach to semantics and pragmatics that proceeds by assigning *semantic values* to sentences, and then reasoning about pragmatics in terms of these semantic values. Classically, the semantic value of a sentence is a *proposition*, constructed as a set of possible worlds. This chapter looks at the application of the same scheme to the semantics of pictures, and to the semantics and pragmatics of narratives such as comics that are constructed as sequences of pictures. Pictures are mapped to propositional semantic values by a geometric method (Section 2). On this basis, an account of the temporal interpretation of pictorial narratives is constructed (Section 3). Another pragmatic aspect of pictorial narratives is anaphora or indexing across pictures in a sequence (Section 4). An argument based on sentences that describe pictures indicates that the semantics for pictures should represent a distinguished viewpoint (Section 5). There is an analogy between ambiguity in language and ambiguous pictures (Section 6). Perceptual semantics for pictures considers the effect of pictures on viewers (Section 7).

2. Possible worlds semantics

This section explains a propositional semantics for pictures, on the basis of a possible worlds construction of propositions. The idea is developed that the information content of a picture, just like the information content of a sentence, is a set of possible worlds.

Suppose we have the background assumption that the world is populated by finitely many solid cubes and octahedrons, and nothing else. Then we are given the sentence in (1), which is to be interpreted as providing information about the world. Understood in this way, it rules out some possibilities, namely those where there are no octahedrons and those where there is only one octahedron. Next we are shown picture (2), which is also to be interpreted as providing information

1

about the world, as if it were an image from a surveillance camera. Much like the sentence, it rules out some possibilities, and allows (or admits) others. For one thing, possibilities where there are no cubes and ones where there is only one cube are ruled out. Other possibilities are also ruled out, such as ones where edges of the two cubes are not aligned.

(1)   There are at least two octahedrons.

(2)
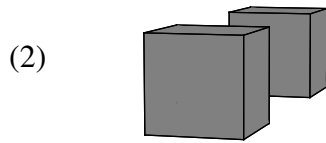
This process of checking possibilities is the basis for the *possible worlds model* of information content.[2] Any informational entity such as a sentence or picture rules out some possibilities, and admits others. The possibilities are called possible worlds, situations, or scenes. It is an easy move to actually identify the semantics of a sentence or picture with the set of possibilities that it admits. This is a set of possible worlds, or proposition. In the semantics of language, the double bracket notation seen in (3) is used as a designation for the propositional "semantic value" of a sentence. In (4) the notation is extended to pictures, simply by putting a picture rather than a sentence inside the brackets.[3]

(3)   ⟦there are at least two octahedrons⟧

(4)   ⟦ ⟧
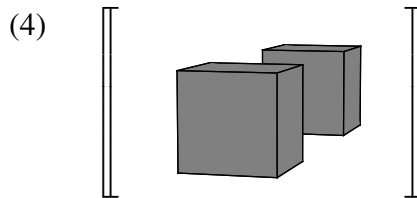
To develop a concrete propositional semantics for pictures, one has to work out a precise characterization of what pictures admit what worlds. The most important approach to this refers to procedures of *geometrical projection* which map three-dimensional scenes to pictures. These have their origin both in theorization about vision, and in geometric methods for creating pictures that were developed in the Renaissance. Essentially, a projection procedure is a mathematical recipe for mapping a three-dimensional scene to a two-dimensional pictorial array. A nice setting for illustrating this is provided by software packages that allow pictures to be created from 3D scenes that are modeled as computational data structures. These pictures can be simple ones like the cube picture, or sophisticated ones like the image shown in (5) of the heroine of the computer-animated film *Sintel* (Blender Institute 2010).

(5)

(6) shows a simple scene data structure, consisting of ordered lists of objects, their geometric type, scale, location, and orientation. Data structure w_1 specifies two cubes. World w_2 is the same, except that the first object has a different orientation. World w_3 is like w_2, but with an additional octahedron.

(6) Possible world w_1

type	scale	translation	rotation
cube	1.0	[0,0,0]	[0,0,0]
cube	1.0	[3,0,0]	[0,0,0]

Possible world w_2

type	scale	translation	rotation
cube	1.0	$[0, 0, 0]$	$[\pi/4, 0, 0]$
cube	1.0	$[3, 0, 0]$	$[0, 0, 0]$

Possible world w_3

| type	scale	translation	rotation
cube	1.0	[0,0,0]	$[\pi/4, 0, 0]$
cube	1.0	[3,0,0]	[0,0,0]
octahedron	1.0	[-3,-2,0]	[0,0,0]

Structures like this can be used as a proxy for possible worlds, so that the proposition that is the content of a sentence or of a picture is a set of such structures.[4] In addition to the objects, a projection procedure involves reference to a *viewpoint*, which for our purposes can be taken to be a location in space together with a vector that locates a *picture plane* that is perpendicular to the vector. A *projection line* is a directed line from the point through the picture plane.

A picture is created by marking points in the picture plane according to a marking rule. For the cube picture, a point in the picture plane is marked in black if the line from the viewpoint through that point intersects an edge of an object before it intersects any other part of an object. Otherwise it is marked in gray if it intersects an object. Otherwise (if the line intersects no object) it is marked in white. This procedure determines a picture in black, gray, and white from a world and a viewpoint, and picture (2) was actually generated computationally from world w_1 and a certain viewpoint.[5] To finish the construction, we say that

the propositional semantic value of a picture p is the set of worlds w such that for some viewpoint v, world w projects geometrically to p.[6]

   An important consequence of this scheme is that one picture admits many worlds. For instance, picture p_2 in (7) admits both world w_2 and world w_3. The reason is that from the viewpoint v for which w_2 projects to p_2, also w_3 projects to p_2, because the octahedron is obscured by the cubes. Picture p_3 is another picture that is consistent with w_3, but this time not with w_2, because there is no octahedron in world w_2.
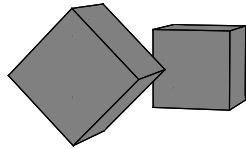
(7)   Picture p_2 Picture p_3
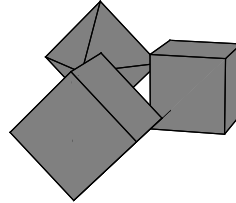
   The fact that a given picture corresponds to many scenes is comparable to the situation in language, where a given sentence is usually true in many possible situations. For instance sentence (8) is true in my current situation, and also true in a situation which differs just in my having an additional chocolate bar in my backpack.

(8)   There is a cat on my lap.

   Different marking rules produce different pictures. The following marking rule results in a "line drawing": mark a point of the picture plane in black if the directed projection line intersects the edge of an object, and otherwise in white. The result can be seen in (9). Independently, the definition of projection lines in terms of the viewpoint can be varied. Orthographic projection uses parallel projection lines perpendicular to the picture plane, rather than lines converging on the viewpoint. Picture (10) is an orthographic version of (2).

(9)

4
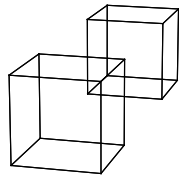
(10)
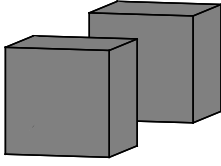
Since the picture depends on the marking rule, the marking rule needs to be treated as a parameter in the assignment of semantic values. The same is true of the geometric characterization of projection lines in terms of the viewpoint. Finally we can also include the model scheme, call it M, which defines a class of possible worlds. The geometric data structures introduced above are a specific construction of M. Writing all of these parameters as a superscript outside the semantic value brackets, $[\![p]\!]^{M,l,m}$ is the propositional semantic value of picture p that is obtained in scheme M, using the definition of projection lines l and marking rule m.

Projection defined in this way gives propositional semantic values for pictures in a manner that is just as precisely defined as the propositional semantic values for sentences that are defined by a formal semantics for a language. We started with the idea that pictures (like sentences) are informational entities that are consistent with some possibilities, and rule others out. Projective semantics for pictures gives a concrete development of this, in that it specifies what possible worlds are consistent with a given picture.

The presentation of a semantics of pictures given here is an idealized and simplified construction that illustrates possible worlds semantics for pictures. It can be extended in many ways. Contemporary models from computer graphics and computer vision take into account factors such as light sources, the reflectance properties of surfaces, and the possibility of multiple light paths. Some extensions are significant for the understanding of artistic style and conventions. (11) shows a picture that is generated by computer from digital camera recordings, which combines information from a strip of viewpoints along the street (Roman, Garg, and Levoy 2004). (12) shows a painting that exploits classical Indian conventions. The woman and fence are projected from a frontal perspective, while the floor of the courtyard is projected from above. While this might appear unsystematic or "incorrect", research in computer graphics shows that projection procedures like this can be systematized to the extent of being implemented in algorithms.

It is worth pointing out that the pictures that are assigned semantic values by the geometric construction are mathematical pictures, not constructions from canvas, wood, and paint or artifacts obtained by drawing with a pencil on paper. To relate the two, one can talk about how physical paintings approximate mathematical ones. An interesting perspective on this is provided by non-realistic rendering in computer graphics, where (in one approach) after a two-dimensional

mathematical picture is rendered geometrically from a three-dimensional scene, transformations are applied to the two-dimensional picture that imitate painting and drawing styles (e.g. Strothotte and Schlechtweg 2002).

(11)

Reproduced by permission from Roman, Garg, and Levoy (2004).

(12)

Painting on panel by Shambhusingh Chaubadar. Reproduced by permission from a private collection.

Summing up, once a projection scheme that maps scenes to pictures is specified, a propositional semantic value for pictures is obtained by inverting the projection. The semantic value of a picture p is the set of worlds that project to p with respect to some viewpoint. This account can be called the *projective model* of the semantic content of pictures. Analyses with this essential character have been discussed for centuries, though often they are combined with theorization about perception and cognition, rather than being presented as a formal semantics for pictures. Sedgwick (1980) is a presentation of the geometry of projection and the information conveyed by pictures that covers also connections to perceptual psychology. Greenberg (2013) is an investigation of the semantic perspective reviewed here. Szeliski (2010) is a textbook presentation of the mathematics of geometric projection from the perspective of computer graphics and computer vision.

3. Temporal progression

Comics and similar artifacts are construed as *sequences* of pictures, corresponding to reading order. Often this order has an interpretation of narrative progression, where the scenes in the described world that verify successive pictures are in temporal succession. For instance the page in (13) from O. Tezuka's *Ode to Kirihito* shows a bathhouse, and in the vertical strip on the left, a man going down some steps, approaching the water, getting in, and rushing out after seeing the dog-like face of the man in the water (the hero Kirihito). The described situations for the four small panels are understood to be in temporal succession. This section and the next one look at this issue of temporal interpretation of picture sequences, and at the problem of *co-reference* across pictures in a sequence.

(13)

Page from Tezuka (1970-1971). ©Tezuka Productions. Used by permission.

　　Panels in classic comics are laid out two-dimensionally on a page. There is a simple scheme based on recursive stacking of panels that determines a linear order based on the layout. It results in a "parse" for a comic page as a labeled tree, which in turn fixes the linear order (Tanaka et al. 2007, Cohn 2013). Suppose we are given the two rectangular pictures that are numbered 28 and 27 in the comic page in (14). The two can be "stacked" horizontally by scaling the vertical dimensions to match, and then assembling them into a larger rectangular image in the order 28,27, following the right-to-left reading order of manga. In the labeled tree (15), this operation of assembly corresponds to the labeled sub-tree $[_H\ 27\ 28]$, where H indicates horizontal stacking. Panels 29 and 30 are stacked horizontally in the same way, forming a larger rectangular area. Next the units $[_H\ 27\ 28]$, $[_H\ 29\ 30]$, and a single panel 31 are stacked *vertically*, this time scaling the horizontal dimensions to match. The label in the tree is V, for vertical stacking. By this means, a page consisting of a number of panels is assembled, in parallel with

construction of an ordered, labeled tree. Linear order for the panels is then read off the order of terminals nodes in the tree. In this case the order is $27 < 28 < 29 < 30 < 31$, indicating the panel-reading order for manga.

(14)

Backround image redrawn from Tanaka (1992). Artist Milka Green.

(15)
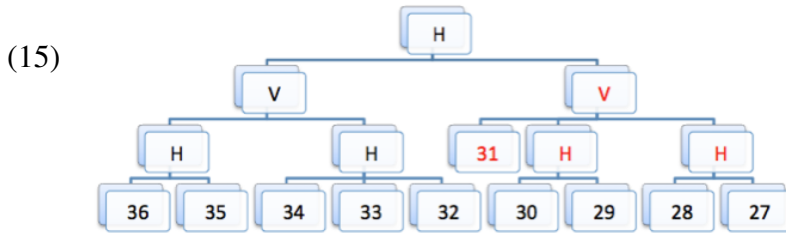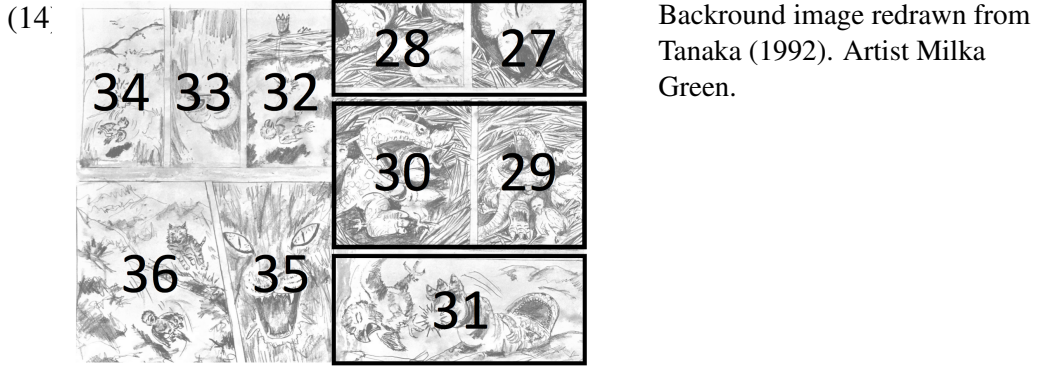
Now we are ready to look at the *interpretation* of a picture sequence. Given the propositional semantics from Section 2, each picture p_i in a sequence $(p_1, ..., p_n)$ has a propositional interpretation $[\![p_i]\!]^{M,l,m}$, which is a set situations.[7] In a world-time line, objects are in different locations at different times. Therefore when we start to consider time, the right notion of situation is a *world at a time*. So we say that a picture is true (or not true) of a pair of world w and a time t. By way of illustration, (17) states approximate semantic conditions for the pictures in the three-element picture sequence in (16). It is understood that a world is temporally extended, and that by adding a time, we identify a "time slice" of a world. The conditions are approximate because the pictures carry additional geometric information about the relative orientation of the cubes.

(16)
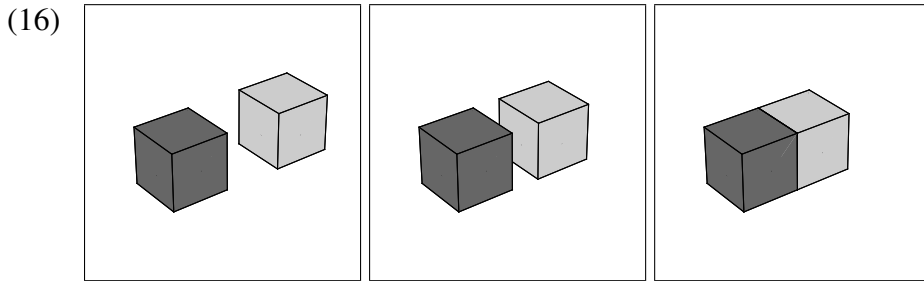
(17) *Approximate semantics for pictures in (16).*

p_1 is true of w, t iff in world w at time t, there is a black cube
at a distance from a white cube that equals the size of the cubes.

p_2 is true of w, t iff in world w at time t, there is a black cube
at a distance from a white cube that equals half the size of the cubes.

p_3 is true of w, t iff in world w at time t, there is a black cube
immediately adjacent to a white cube.

The project now is to combine the propositions in (17) into an interpretation for the sequence (p_1, p_2, p_3), in a way that requires temporal progression. The semantic value should be a set of objects of some kind. The individual objects in the set should be temporally extended, to represent passage of time in the described situations. Given the assumption that a world is temporally extended, worlds can be used as the elements of the semantic value. But in order to give a recursive definition of the interpretation of a sequence, it is helpful to include also a time parameter, which is understood as relating to the time at which the last picture in the sequence is satisfied. This allows time to be moved forward with each successive picture. (18) defines an interpretation for picture sequences in a way that requires invariant temporal progression. The definition is stated as a recursive definition of a picture sequence being true of a world w and a time t. The latter records a time at which the last picture in the sequence is satisfied. If Σ is true of the world-time pair w, t, time t precedes time t', and p is true as a picture of w, t', then the longer sequence Σp is true of w, t'. Finally a sequence Σ is true *simpliciter* of w iff for some time t, Σ is true of w, t.

(18) *Invariant progression*

Base Let p be a picture. The unit sequence (p) is true of w, t iff p is true of w, t as a picture.

Recursion Let p be a picture and Σ be a picture sequence. Σp is true of w, t' iff there is a t such that the following conditions are satisfied.
(i) Σ is true of w, t.
(ii) p is true as a picture of w, t'.
(iii) $t < t'$.

A picture sequence Σ is true of w iff for some t, Σ is true of w, t.

Is time really understood as progressing invariantly from picture to picture in works such as comics? In discussion of temporal progression in natural language, it is often maintained that there are sentences in narratives that do not "move time forward" (Kamp and Rohrer 1983, Hinrichs 1986, Kamp and Reyle 1993). There is a good case for this with stative sentences.[8] The natural interpretation of example (19), where the second sentence is stative, requires the door to be ajar at the time of arrival.

(19) Marlowe arrived at his apartment. The front door was ajar.

McCloud (1994) pointed out types of panel sequences in comics where "there is little feeling of temporal movement." These frequently involve panels which, informally, have a stative character. The bathhouse sequence in (13) is an example. The long panel on the right shows a scene with a bathhouse below a mountain. It is presumed that the configuration of the bathhouse and the mountain are the same during the events depicted by the strip on the left. The sequence (20) in Masashi Tanaka's *Gon*, is a different case. The top panel shows a bobcat on its back. The second shows some other animals flying above. One could claim that the times for the two panels are identical, because the second panel shows what the bobcat sees in the first.
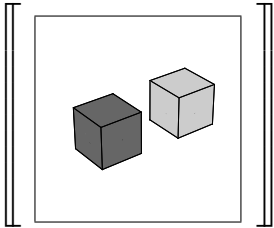
(20)

Redrawn from Tanaka (1992).
Artist Milka Green.

These data suggest the possibility that sequences involving a "stative" panel have an interpretation of temporal overlap, rather than temporal succession. This could be expressed in a rule that moves time forward for non-stative pictures, but not for stative ones. This is similar to the aspectually sensitive rules for temporal succession that are hypothesized to cover linguistic data such as (19).[9] Using aspectually sensitive principles in the account of linguistic narratives is regarded as well-motivated. There are a couple of problems with extending this to pictorial narratives. First, the formal analysis of aspectual distinctions such as stativity is usually developed in terms of patterns of truth with respect to points and intervals of time (Dowty 1979). One property of statives is that they can be true at points of time, and can be true at every point in a continuous interval. This condition is intuitively satisfied for statives sentences such as those in (21). It is not satisfied for the "accomplishment" sentence (22a), because building a cabin takes time. And it is not satisfied for the "achievement" sentence (22b), because one can't

10

reach the summit "again and again" at each point in a continuous interval. By the way, while such judgments have some relevance, semantic theories of aspectual type proceed from formalized accounts of lexical and compositional semantics, where semantic definitions of properties such as stativity are expressed in logic and mathematics (Taylor 1977, Dowty 1979, Krifka 1989, Abusch 2005).

(21) a. Jack is asleep.
    b. The center of mass of the building is at point x.

(22) a. Jack built a cabin.
    b. Jack reached the summit.

The problem for the pictorial case is that by these criteria *all* semantic contents for pictures come out as stative. Consider the proposition indicated on the left in (23), obtained by taking the semantic value of the picture inside the brackets. Suppose that in world w there is a scene with a black and a white cube that projects to the picture with respect to viewpoint v at time point t. If the cubes are static in an interval i that contains t, then the scene projects to the same picture with respect to v at any time point t' in i. So the proposition (23a) is true of w, t' for every point t' in the interval i. And if one extends the truth definition to include truth with respect to intervals, then the proposition is true with respect to every subinterval of i. This pattern of truth is characteristic of stative propositions.

(23)   a.   b.

Image on right from Tatsumi (2008). ©Yoshihiro Tatsumi 2008, 2013. Used by permission.

The same reasoning applies even for propositions expressed by pictures that are intuitively dynamic, such as proposition (23b), formed as the semantic value of a manga panel showing three boys playing basketball. This by the way also applies in the linguistic case, where sentence (24) refers to a comet that we presume to be moving, but the proposition it denotes is linguistically stative.

(24)   The center of mass of the comet is at point x.

The reasoning above indicates that the propositional semantic content of a picture is in all cases stative. This makes it impossible to state an interpretation rule for picture sequences that refers to a semantically distinguished subset of "stative" propositions (Abusch 2014).[10] This is different than the linguistic case, where a distinction between stative and non-stative sentence meanings is well-supported.

There is also a basic problem with the hypothesis that the interpretation of

sequence (13) involves temporal overlap. In the picture on the right, some vapor is depicted above the spring. We do not assume that the vapor is motionless over the time taken up by the actions inside the bathhouse.

Summing up, in the semantics of picture sequences it is *not* possible to trigger an interpretation rule from a semantic property of stativity of propositions. One option is to stick with some kind of invariant rule. This could be the invariant progression rule (18). An alternative is a principle of invariant non-regression, which is obtained by replacing $t < t'$ in (iii) of (18) with $\neg t' < t$. This allows the verifying time for two successive frames to be identical in some cases, or if intervals are used instead of times, to overlap. This is attractive for sequences like (20).

4. Indexing

Example (25) is the three-panel sequence from Section 3 that depicts two cubes moving together while maintaining alignment. In one significant dimension, this description is stronger than the semantic content that is delivered by the semantics from Sections 2 and 3. Including temporal succession and assuming a propositional semantics, the literal content of (25) can be described approximately like this. The proposition corresponding to this comic is true in a world w only if there are times t_1, t_2, t_3 such that that $t_1 < t_2 < t_3$, where at t_1 in w, there is a black cube at a distance from a white cube that equals the size of the cubes, at t_2 in w, there is a black cube at a distance from a white cube that equals half the size of the cubes, and at t_3 in w, there is a black cube immediately adjacent to a white cube. Missing in this is the information that the black cube in the first picture is the same as the black cube in the second picture and the black cube in the third picture. In the same way, the literal content of (13) does not include the information that the dog-like man in the third small picture is the same as the dog-like man in the fourth small picture. Certainly, someone who reads the sequence makes the inference of identity, and the author intends for readers to do so.

(25)
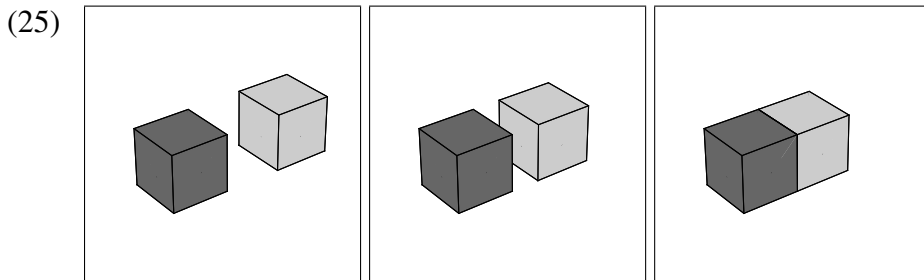
In accounts of the syntax and semantics of natural languages, it is common to assume that syntactic representations and representations of discourse contents include encoding of co-reference (Karttunen 1969, Chomsky 1980, Kamp 1981,

Heim 1982, Pollard and Sag 1994). Indices can be included in syntax, as in (26a). Or the identities can be expressed at a level of discourse representation, using discourse referents, which are syntactic objects semantically similar to variables. In part, indexing is constrained syntactically, so that (26b) has only a co-indexed reading—this is part of the motivation for including indices in syntax or at the interface between syntax and semantics.

(26)a. $[_{DP}$ a woman$]_1$ moved toward $[_{DP}$ a rabbit$]_2$. $[_{DP}$ she$]_1$ picked $[_{DP}$ it$]_2$ up.
   b. $[_{DP}$ a woman$]_1$ pinched $[_{DP}$ herself$]_1$.

(27) is version of (26a) where all the nominals are existentially quantified, because they are indefinite descriptions. Semantically, this example is parallel to the pictorial example (25), with all of the represented objects existentially quantified. There seems to be no "reading" where co-reference is intended. This could be because a representation with co-indexing is ruled out for indefinites by a semantic requirement that indefinites have new indices.[11] Or it could be because a co-indexed reading is blocked by the more specific alternative (26a).

(27) $[_{DP}$ a woman$]$ moved toward $[_{DP}$ a rabbit$]$. $[_{DP}$ a woman$]$ picked up $[_{DP}$ a rabbit$]$.

For the the pictorial case, it falls out of the semantics that depicted objects are in effect existentially quantified. There is no definiteness distinction, and there is nothing that enforces identity of objects depicted in different pictures. Nevertheless there is a pragmatic variety of co-reference, in that readers infer identity between depicted objects, and authors intend for them to do so. One strategy for addressing this situation is to introduce representational objects similar to indices, and to interpret them using the same techniques that are applied to indices in language. This has to be done in a way that constrains what objects in the described situation can correspond to a given index. In the linguistic case, this is done by referring to the syntactic constituents and their semantic interpretations. In a semantics for sentence (26b), a value for index 1 gets constrained to be a woman in the model, using an interpretation rule that refers to the fact that *a woman* is a DP with index 1. It is difficult to mimic such interpretation rules in the pictorial case, since on the analysis from Section 2, pictures do not have syntactic sub-constituents that are significant for semantics.

However, they do have unstructured parts that can be placed in correspondence with the scene. Abusch (2012) suggested constructing pictorial indices as areas of pictures. An area is a part of the plane, such as a circular area defined by a center and radius. (28) is a comic of a cone moving in front of a torus, while the bottom of a tube bends up.[12] Each picture has been marked with a circle, which functions as an index for the cone. The projection lines that relate scenes to the picture also establish a correspondence between areas of the picture and objects in the scene, so that areas can be used syntactically as indices. An object corresponding to a given area is some object which meets the constraint that lines from the viewpoint

through the area intersect that object before it intersects any other object.

(28)

The basic semantics of (28) is consistent with worlds where a single cone moves in front of a torus. It is also consistent with worlds where the cone of the first picture moves out of view, and another cone moves into view. To infer identity between the cones is to eliminate worlds of the second kind. This is done by adding to the discourse representation a syntactic predication of identity between the two indices, serving the same function as co-indexing in linguistic representations.

Formally, an enriched pictorial sequence is a linear sequence consisting of pictures, geometrical areas in pictures, and syntactic identities. In (29), the areas a and b are circles, defined by center and radius. An area is understood to be an area in the preceding picture. The formula $1 = 2$ uses a recency convention, where 1 is the most recently introduced picture area, and 2 is the previously introduced area. In this case $1 = 2$ functions as a syntactic identity between the discourse referents for the two cones.

(29)
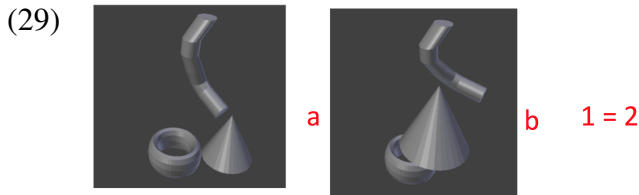
Interpreting enriched sequences requires machinery to keep track of referents. The starting point is the recursive definition from Section 3 of a world and a time satisfying a picture sequence. To this is added a sequence $\mathcal{O}$ which records entities that serve as referents for indices. In addition it is necessary to keep track of the viewpoint that was used for the last picture in the sequence, because this is used in interpreting new indices. So, the semantic objects manipulated by the definition are tuples $w, j, v, \mathcal{O}$, where w is a world, $\mathcal{O}$ is a sequence of entities, and j, v are the time and viewpoint that were used for projecting w to the last picture in the pictorial sequence.

(30) and (31) are two representative semantic rules. (30) interprets a sequence of the form $\mathcal{D}a$ which has an area a at the end. The idea is that referring to the time, the viewpoint, and the area that is serving as an index, an entity in the world is picked out by following projection lines from the viewpoint through the area. This entity o is recorded by putting it at the front of a new entity sequence $o\mathcal{O}$. Because the rule refers to entities in the world w, it requires worlds in which entities are

14

individuated. This is so for the geometrically constructed scenes discussed in Section 2, where the specification of a scene includes a list of geometric entities such as cubes. In order to interpret equalities, entities in fact need to be identified across different times in a world.

(30) $w, j, v, \mathcal{O}$ satisfies $\mathcal{D}$
        a is an area
        For any directed line d from v through area a, the minimal point
        on d that is on the surface of an object in w, j is on the surface of
        object o in w, j.
        ──────────────────────────────
        Then $w, j, v, o\mathcal{O}$ satisfies $\mathcal{D}a$.

(31) is the rule for an enriched pictorial narrative like the one in (29) that has an equality between indices at the end. To check the equality $m = n$, one simply looks up the mth and nth entities in the sequence $\mathcal{O}$ and checks equality between them. For the equality $1 = 2$ to hold, the first object in $\mathcal{O}$ has to be the same as the second object in $\mathcal{O}$.

(31) $w, j, v, \mathcal{O}$ satisfies $\mathcal{D}$
        $\mathcal{O}[m] = \mathcal{O}[n]$
        ──────────────────────────────
        Then $w, j, v, \mathcal{O}$ satisfies $\mathcal{D}(m = n)$.

This kind of enriched pictorial narrative, where invisible indices are inserted into the picture sequence, is parallel to linguistic representations where inaudible indices are included in the syntactic tree. The scheme of dynamic interpretation is also similar to what is used in the linguistic case. In particular it is similar to Decker (2012), where a recency convention for indices is used. There also are substantial differences. Pictorial indices are constrained by content using a geometric condition, rather than using indices that are located on particular nodes in a syntactic tree. The dimension of novelty/familiarity comes out differently: all indices are novel when they are introduced as areas in a picture, and are familiar in equality conditions. And there is no counterpart of a morphosyntactic definiteness distinction. From the standpoint of linguistic syntax and semantics, the striking property of this system is that it can be characterized as entirely post-semantic, with indices and equalities between them having the status of enrichments of basic semantic content.

Does indexing in natural languages ever follow the pictorial model? One candidate is ordinary nominal reference in languages without articles and without post-nominal definiteness marking. Such languages are common, and include Chinese, Serbo-Croatian, and Warlpiri. In syntactic research, it is debated whether languages without articles project DP, or have bare NPs. Boscovic (2008) and Despic (2013) give syntactic arguments for the bare NP hypothesis. But even assuming this syntactic analysis, there is still the possibility that a definiteness distinction is created at the syntax-semantics interface, so that nominals can be

either definite or indefinite in literal semantics. Bittner and Hale (1995) make this proposal for Warlpiri.

Pictorial indexing suggests the possibility that bare NPs are primitively indefinite, and can be in effect converted to definites by the addition of an equality among indices. In the pictorial case, this process is post-semantic. This is probably not right for the linguistic case, because there are good arguments for interactions between syntax and definiteness. In Mandarin Chinese, articleless NPs at the left periphery of the clause are interpreted as definite (e.g. Cheng and Sybesma 1999). This interaction between syntax and definiteness indicates that definiteness needs to be present syntactically, and therefore pre-semantically. A tricky approach is to posit that Mandarin NPs are basically indefinite, and that the co-indexing that creates a definite is contributed by the features and projections that trigger movement to the left periphery. In this case NPs are primitively indefinite in Mandarin, but indexing is still pre-semantic.

In English, there is no definiteness distinction for verbs. Where verbs introduce event variables, that event variable is treated as existentially quantified at the clausal level (e.g. Beck and von Stechow 2014). But optionally, events described by verbs can be construed as identical with previously mentioned events. For instance in (32), it is understood that the disassembly event introduced by the second sentence is the same as the disassembly event introduced by the first sentence.

(32) An airplane engine was disassembled and rebuilt. Jack disassembled it.

5. Picture descriptions and centered content

Not only do pictures have semantic content, their semantic content can be described in terms of natural language. The sentences in (33) illustrate locutions that describe the content of pictures. In the semantics of English, such sentences should be characterized as true or false depending on properties of the picture that is mentioned. A plausible basic setup is that a picture is an entity, something of semantic type e. The semantic construction from Section 2 maps such entities to propositional contents. Suppose we assume also that sentences have propositional semantic contents. Then it is possible to state a semantic rule for the construction in (33a). We say that a sentence of the form "in x, ϕ" is true if and only if any world that is an element of the propositional content of x is also an element of the propositional content of ϕ. Equivalently, "in x, ϕ" is true if and only if the propositional content of x entails the propositional content of ϕ. This is recorded in (34).

(33) a. In the picture, there is a man on a couch.
b. The picture shows a man on a couch.

(34) A sentence of the form "in x, ϕ" is true if and only if the propositional content of medium x entails the propositional content of sentence ϕ.

Some good consequences follow immediately. A basic semantics for English already defines a propositional content for (35a) that entails the proposition contributed by (35b). It then falls out from the semantic rule and some simple logic that sentence (33a) entails sentence (35c).

(35) a. There is a man on a couch.
     b. There is a person on a couch.
   c. In the picture, there is a person on a couch.

This kind of analysis of sentences describing pictures was given by Ross (1997). He pointed out that the constructions in (33) can be used also with movies, novels, and comics—see (36). In his terminology, the content-bearing object x is a *medium*. Ross proposed a formal language from modal logic that includes a family of sentential operators $[x]$, which are used in logical formulas of the form $[x]\phi$. The operator $[x]$ is defined only if x is a medium. (37) shows the formalization of an English sentence in this language.

(36) a. In the video clip, a man falls off a couch.
     b. In the novel, a woman falls in love with a nobleman.
     c. In the comic, a man develops a disease that gives him a dog-like face.

(37) a. In one picture, there is a man on a couch.
     b. $\exists x.\mathbf{picture}(x) \wedge [x]\exists y\exists z[\mathbf{man}(y) \wedge \mathbf{couch}(z) \wedge \mathbf{on}(y,z)]$

Something to notice about the approach is that the clause (35a) has exactly the same semantics when used in the picture-description (33a) as otherwise. There is no need to give a special semantics (perhaps a geometric one) for sentences that are used in picture-descriptions. Everything that is specific to pictures has been packed into the specification of semantic values for pictures. All of this counts as motivation for giving pictures a propositional semantics: it allows English sentences that describe the content of pictures to be integrated smoothly into the semantics of English.

A further issue is the phenomenon of *perspectival* descriptions of pictures. Ross (1997: Ch. 4) pointed out that the two pictures (38a,b) have a rather similar propositional content, along the lines of 'there is a white ball and a black ball'. Yet sentence (39) is judged to be true when the picture it refers to is (38a), and false when the picture it refers to is (38b). This is a problem if the pictures have identical propositional contents, because in the compositional rule (34), pictures enter into the semantic content of sentences describing them via their propositional contents, and this is assumed in the argument to be the same for the two pictures.[13]

(38) a. b.
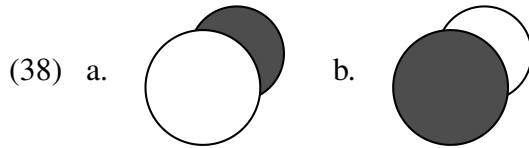
(39) In the picture, there is a white ball in front of a black ball.

Ross (1997) pointed out a connection between perspectival picture descriptions and the phenomenon of de se readings in attitude descriptions. There is a family of examples and arguments that indicate a special status for reference to the attitude-bearer in linguistic descriptions of propositional attitudes.[14] For instance Perry (1977) tells the story of an amnesiac Lingens who is lost in the Stanford library, and reads a biography of himself, without realizing that he is the subject. Suppose that based on online reading about upcoming conferences, Lingens forms the expectation described in (40a), again without realizing the identity. Sentence (40a) is true in this scenario. Yet sentence (40b) is false. This version has a null subject, known linguistically as a PRO subject. In infinitival attitude complements, PRO subjects are observed to have a special status related to self-knowledge.[15].

(40) a. Lingens expects Lingens to be the subject of an interdisciplinary conference.
b. Lingens expects to be the subject of an interdisciplinary conference.

Lewis (1979) dubbed such readings *de se* readings, and proposed an analysis where complement sentences in attitude descriptions contribute properties rather than propositions. The relevant property in the de se sentence (40b) is the property of being an individual x_{self} such that there will be an interdisciplinary conference about x_{self}. Speaking informally, Lingens does not self-attribute this property, accounting for the fact that (40b) is false. He does self-attribute the property of being an x_{self} such that there will be an interdisciplinary conference about the individual that x_{self} has read about under the name "Lingens". This is a basis for an explanation of the fact that (40a) is true.

Lewis's semantics is perspectival in that it works with a model of information that includes identification of an attitude holder in a possible world—this is the x_{self} in the previous paragraph. Now the propositional semantics in Section 2 also worked with a special element of the world, namely the viewpoint. It is an easy move to bind the viewpoint to form a property, rather than existentially quantifying it, as was done in Section 2. This leads to the hypothesis that the semantic content of a picture is a *property of viewpoints*. Semantically this can be constructed as a relation between a viewpoint v, world w and, taking into account Section 3, a time t. This results in semantic values of pictures following the schema (41). The difference from Section 2 is that the viewpoint is bound by the lambda operator, rather than being existentially quantified.

(41) $[\![p]\!]^{M,l} = \lambda v.\lambda w.\lambda t \left[\begin{array}{l} w \text{ at } t \text{ projects to } p \text{ via projection method } l \\ \text{and marking rule } m \text{ using viewpoint } v \end{array} \right]$

Changing the semantic value of pictures from propositions (sets of worlds) to relations between viewpoints and worlds allows the viewpoint to be taken into account in characterizing the semantics of the sentences in (33). This should help, because phrases such as "in front of" are clearly perspective-sensitive. But what exactly is the semantic rule that replaces (34)? Ross (1997:5.3) and Blumson (2010) developed accounts that follow Lewis closely. Sentence (42) contributes a property of viewpoints rather than a proposition. Independently, picture contents are the properties of viewpoints that are defined by a semantics like (41). Then it is possible to revise (34) to work with properties rather than propositions, much as Lewis stated a semantics for the verb *believe* that is based on properties rather than propositions. This rule is given in (43).

(42) There is a white ball in front of a black ball.

(43) A sentence of the form "in x, ϕ" is true if and only if every world-time-viewpoint tuple $\langle w, t, v \rangle$ in the semantic value of medium x is an element of the viewpoint-centered semantic value of sentence ϕ.

What is responsible for the viewpoint-sensitivity of the clause (42)? Ross (1997: Section 5.3) pointed out that substituting *near* for *in front of* results in a sentence (44) that is not viewpoint-sensitive. This suggests that viewpoint-sensitivity comes from a subset of lexical items such as *in front of*. It is notable that the pragmatic interpretation of (42) requires reference to a contextually supplied viewpoint, which can be established by the orientation of the speaker and addressee. Technically, we can say that some lexical items contribute a viewpoint parameter, which has a special pragmatic interpretation when it is free, and which can also be bound in constructions such as (33).[16]

(44) There is a white ball near a black ball.

Ross (1997) and Blumson (2010) both stress the parallelism between this analysis of viewpoint-sensitive descriptions of pictures and Lewis's theory of de se reference in attitude descriptions. Should the two be identified? Suppose properties of individuals as used by Lewis were identified with properties of viewpoints as used in the semantics of pictures and the semantics of picture descriptions. We know that PRO in the syntax-semantics interface has a special role as the expression of de se readings by verbs with infinitival complements. And *in front of* has the special role of contributing a viewpoint parameter that is used in the semantics of picture descriptions. If de se properties and viewpoint properties are the same thing at the syntax-semantics interface, then these should be interchangeable. For instance, in (45) the viewpoint parameter should get bound in the same way as PRO gets bound in (45b). This would presumably result in a reading that defines

19

in front in terms of Keisha's future location. There is no such reading.

(45) a. Keisha expects Justin to sleep in front of a boulder.
    b. Keisha expects PRO to sleep in a motel.

In general, the lexical and constructional sources of viewpoints seem to be independent of the lexical and constructional sources of *de se*. This indicates that the viewpoint parameter should *not* be identified with the de se parameter at the syntax-semantics interface.[17]

The interpretation of indices in Section 4 actually requires viewpoint-centered contents for pictures. The reason is that the tuple $w, j, v, \mathcal{O}$ memorizes the viewpoint for the last picture, in order to use the viewpoint in interpreting indices. To integrate this framework with a semantics for pictures, it is necessary to use viewpoint-centered semantic values for pictures. This can be seen in the semantic rule (31) for a narrative that ends with a picture. The second premise in the rule refers to a world-time-viewpoint tuple w, k, u that satisfies the picture p that comes at the end of picture sequence $\mathcal{D}p$ in the conclusion. It has to be assumed that pictures have a viewpoint-centered semantics to check the second premise.

(46) $w, j, v, \mathcal{O}$ satisfies $\mathcal{D}$
        w, k, u satisfies picture p
        $\underline{j < k \qquad\qquad\qquad\qquad}$
        Then $w, k, u, \mathcal{O}$ satisfies $\mathcal{D}p$.

Summing up, this section has reviewed a semantic analysis of constructions like *in the picture*. The analysis has to assume that a picture contributes a viewpoint-centered proposition. So these data and analysis motivate a revision of the semantics from Section 2. The semantic value of a picture is a viewpoint-centered proposition, rather than a proposition. Technically, it is a relation between worlds, times, and viewpoints.
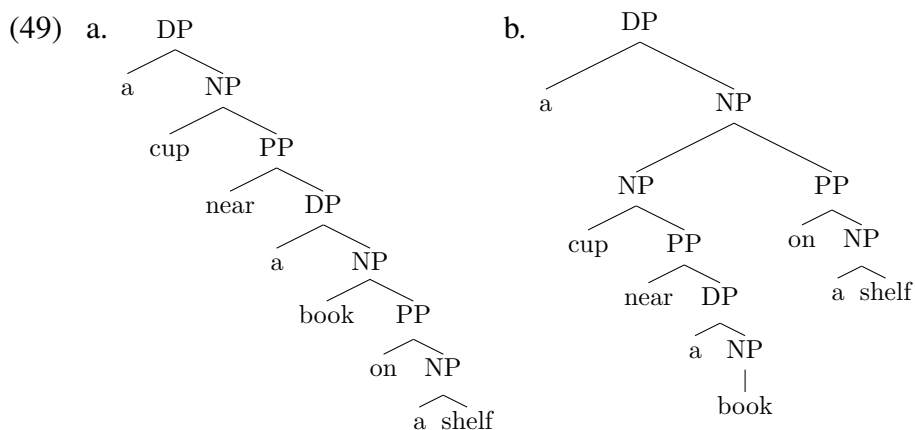
6. Ambiguity and strengthened content

Section 2 mentioned that a picture showing two cubes is true of some situations containing two cubes and nothing else, and also true of some situations containing two cubes and a third object. In general, a picture does not semantically evoke a unique described world. Instead, there are many possible worlds that are consistent with the semantic content of a given picture. This kind of semantic multiplicity is known as *non-specificity*, and it is distinguished from ambiguity (Zwicky and Sadock 1975). To take a linguistic example, sentence (47) is not ambiguous between more specific "readings" of the sort exemplified by (47b) and (47c). It is merely non-specific among situations where Keisha is in a car, situations where Keisha is in an airplane, and other alternatives.

(47) a. Keisha is in a vehicle.
    b. Keisha is in a car.
    c. Keisha is in an airplane.

On the other hand sentence (48a) has a different kind of multiplicity, which amounts to an ambiguity between one reading that entails that there was a book on a shelf, and another reading that entails that there was a cup on a shelf. On the first reading, the book is on a shelf, while the cup might be on a nearby table and not on a shelf. On the second reading, the cup is on a shelf, and the book might be suspended from the ceiling in a position that is near the cup. This is usually attributed to an ambiguity in syntactic structure—the reading that entails (48b) contains a nominal phrase with a tree shape along the lines of (49a), and the reading that entails (48c) has a tree shape (49b).

(48) a. A cup near a book on a shelf spontaneously shattered.
    b. There was a book on a shelf.
    c. There was a cup on a shelf.

(49) a.

```
        DP
       /  \
      a    NP
          /  \
        cup   PP
             /  \
          near   DP
                /  \
               a    NP
                   /  \
                book   PP
                      /  \
                    on    NP
                         /  \
                        a   shelf
```

b.

```
          DP
         /  \
        a    NP
            /    \
          NP      PP
         /  \    /  \
       cup   PP on    NP
            /  \     /  \
         near   DP  a   shelf
               /  \
              a    NP
                   |
                 book
```

Linguistic accounts maintain that the objects that are mapped to meanings are structured objects like the trees in (49), rather than mere strings.[18] Independently, trees can be mapped to word strings, or if we include phonology, to phoneme strings. A syntactically ambiguous word string or phoneme string is one which can be mapped from two or more syntactic structures. If these syntactic structures semantically contribute different propositions, then the word string is both syntactically and semantically ambiguous.

It is attractive to try to draw a parallel between this characterization of ambiguity in language and the analysis of "ambiguous" pictures such as the Necker cube (Necker 1832). The Necker picture (50) is experienced as having a shifting geometrical interpretation. For the viewer, the vertical line above the arrow can "pop out", amounting to an interpretation where the cube edge corresponding to the line is closer to the viewpoint than the rest of the cube. Or it can "pop in", corresponding to an interpretation where the cube edge is further from the viewpoint

than the rest of the cube.

(50)
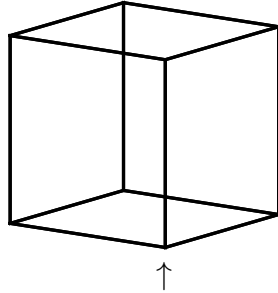
The drawing (50) can be produced from a scene with a cube using orthographic projection, and a marking rule that creates black marks for projection lines that intersect edges of a geometric figure, and white marks for other projection lines. In the notation from Section 2, the semantic value is $[\![(50)]\!]^{M,o,b}$, where o is orthographic projection and b is the black/white marking rule. According to Section 2, $[\![(50)]\!]^{M,o,b}$ is a set of worlds, because the viewpoint is existentially quantified in the construction. With this semantics, there is seemingly no sense in asking whether the viewpoint (or the picture plane) is nearer or further from a particular edge in an element of the denoted proposition. And in a single world containing just a cube, picture (50) can be produced from viewpoints oriented toward each of the eight edges of the cube, so it is not possible to refer to orientation of the cube with respect to a unique viewpoint. In the construction from Section 5, viewpoints are encoded in the semantic value. Specifically, $[\![(50)]\!]^{M,o,b}$ is a set of pairs of worlds and viewpoints.[19] For one pair $\langle w, v \rangle$ in the semantic value, it is possible to use projection lines to track back from the line in picture (50) to an edge of a cube in w, and to check whether that edge is closer to v than the rest of the cube. However it turns out that the same picture (50) is produced from pairs $\langle w, v \rangle$ where in w the cube-edge corresponding to the line above the arrow is closer to v than the rest of the cube, and from pairs $\langle w', v' \rangle$ where in w' the edge corresponding to the line above the arrow is further from v than the rest of the cube. Therefore in the centered proposition $[\![(50)]\!]^{M,o,b}$ the "close edge" pairs $\langle w, v \rangle$ are mixed together with the "far edge" pairs $\langle w', v' \rangle$. Disappointingly, in this semantic model, the Necker multiplicity comes out as non-specificity rather than ambiguity.

If we want to avoid this conclusion, looking at the linguistic case suggests the strategy of adding more structure to the object inside the semantic value brackets. Huffman (1971), Clowes (1971), and Waltz (1972) devised syntax for line drawings that includes $+/-$ features and directional arrow features on line segments. Plus features mark lines that are projected from edges where faces meet at a convex angle with respect to the viewpoint. Minus features mark lines that are projected from edges where faces meet at a concave angle with respect to the

22

viewpoint. Arrows are found on edges that have the body of the object on one side and not on the other, with the body on the right along the alignment of the arrow. In this syntax two readings of Necker picture are annotated as in (51). The picture structure on the left has the line above the arrow marked with a plus sign, marking a convex solid angle that "pops out". The picture structure on the right has the line above the arrow marked with a minus sign, marking a concave solid angle that "pops in".

(51)
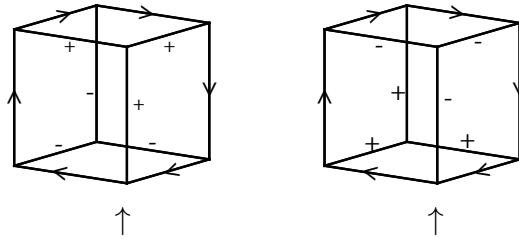
Huffman-Clowes-Waltz features need to be incorporated in the marking rule, so that syntactic pictures include features.[20] We should then treat the "surface picture" (50) as being mapped from either of the picture structures in (51), in the same way as the word string (48) is mapped from either of the trees in (49). This produces a perfect analogy with the linguistic case.

The analogy extends to the problem of finding a structured representation of a surface word-string or a surface picture. Waltz based a pictorial parsing procedure on the observation that the possible labelings are constrained locally. (52) lists the possible local shapes of junctions with three incoming edges. Any line segment must be labeled consistently, with constraints coming from the two vertices that terminate the line segment.

(52)

23
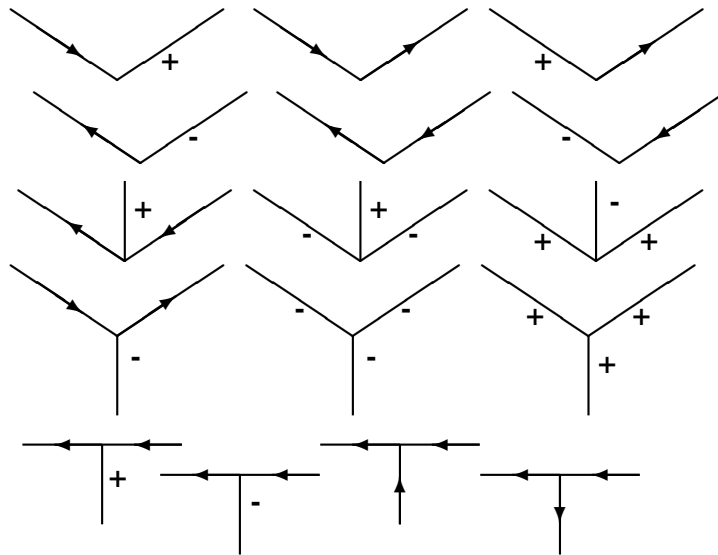
In this scheme, parsing is the problem of finding a structured global representation that satisfies local constraints. This is formally similar to the parsing problem in natural language, where the lexicon of words and a grammar (such as a context free grammar) specify permissible local configurations, and the parsing procedure must find a global representation that satisfies the local constraints.

As an alternative, we could accept the conclusion that the Necker picture is non-specific rather than ambiguous. On this story, picture (50) after all has a single reading, with a single weak semantics that is compatible with both near-edge models $\langle w, v \rangle$, and far-edge models $\langle w', v' \rangle$. Greenberg (2011) suggests accepting this conclusion. This theory needs to explain what to make of the intuition that cognitively, the picture has two interpretations, and that a human viewer can shift temporally between attending to one interpretation and attending to the other. One could say that the two interpretations that we experience are obtained by adding information to the literal semantic content of the picture. The extra information for one enriched meaning excludes "close edge" pairs $\langle w, v \rangle$, and for the other enriched meaning excludes "far edge" pairs.

To an extent, the issue of strengthened readings comes up in natural language semantics. Does *boy* as used in (53) have a weak basic meaning that has both human boys and statues of boys in its extension? Or does it have a basic meaning 'human boy' that can be mapped by some process to a meaning that includes statues of boys in its extension? If we assume a weak basic lexical meaning for *boy*, then the "real boy" reading that is experienced for (53a) must be considered an enrichment of the basic semantics. This kind of analysis is advocated in some literature on lexical semantics (Carston 2002).

(53) a. There is a boy in the entrance.
   b. The design consultants are installing a boy in the entrance.

It is harder to apply the weak strategy to apparent ambiguities arising from combinatory possibilities in syntax, such as the ambiguity in (48). The problem is that there appears to be no way of getting to a weak disjunctive meaning, other than representing the different syntactic readings, computing the propositions they denote, and taking the disjunction of those propositions. And while semantic ambiguity is part of the motivation for syntactic structure, it is not the only motivation—syntactic structure is motivated internally to syntax. The Necker case is different. Notably, a means of getting to a weak meaning is at hand—namely the centered projective semantics from Section 5, without the addition of Waltz features.

Suppose that we consider well-motivated the solution to the Necker problem using Waltz features in the syntax of pictures. This would have the important consequence of providing an additional argument for a viewpoint-centered semantics for pictures. The syntactic-semantic solution to the Necker problem using Waltz features seems to work only with centered propositions, where viewpoints are part

of the encoding of semantic values.

   This discussion brings up the pervasive phenomenon of *strengthened content* for pictures. When we look at pictures, we ordinarily make assumptions which go beyond the semantic content that is supplied by the projection theory. For instance the picture repeated in (54) is understood as depicting three boys who are about the same size and are near each other. But in the projection model, the left boy might be three times as tall as the others, and correspondingly far away. Or the middle figure might be an acrylic statue of a boy, rather than a human boy. Readers of Tatsumi's manga *A Drifting Life*, from which this panel is redrawn, make assumptions that rule out these outlandish possibilities. That is, they infer a content that is stronger than the literal one.

(54)
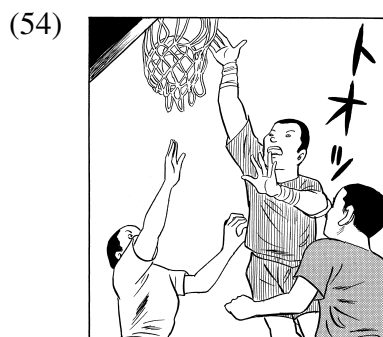
Panel from Tatsumi (2008).
ⓒYoshihiro Tatsumi 2008, 2013.
Used by permission.

7. Transparency and perceptual semantics

Under some conditions, the visual experience of viewing a picture can be perceptually equivalent to viewing one of the scenes that is compatible with the content of the picture. In visiting a building with trompe-l'œil features we may be fooled into thinking that there are elaborate architectural features at the corner of a hall, rather than painting on a smooth curved surface. Looking through a peephole in the side of a box at a painted surface on the opposite side, we may think we are looking into an adjacent furnished room. This property of perceptual equivalence between viewing a picture and viewing the subject of the picture is known as *transparency*. The experience of viewing a picture may be transparent, in that for the viewer, it is just like viewing the depicted scene directly.[21]

   Connected to transparency is the idea of using perceptual-cognitive states or relations of perceptual similarity as a basis for the semantics of pictures, and/or for a characterization of what pictures depict. Peacocke (1987) gives a two-step explanation of the notion of a picture depicting a Φ, for instance depicting a suitcase or depicting a cat. The basis is a notion of perceptual similarity between two regions of an agent's visual field on different occasions. When an agent is viewing a picture of a cat, it is claimed, there is a region of the agent's visual field that is

similar in shape (or other perceptual qualities) to a region of the agent's visual field in another possible situation where the agent is viewing an actual cat. (55) defines a notion of an object p (such as a picture-part) being F-related to a concept Φ (such as the concept 'cat'). A picture-part p is cat-related for agent x if and only if p as presented to x results in a field-region that is similar to a field-region that could result from presenting x with a real cat. The letter F in F-related indicates the visual field of the agent.

(55) p is F-related to concept Φ for agent x if and only if for x, p as presented in x's visual field is similar in shape to a visual-field region in which a Φ could be presented.

Depiction is then defined as intended F-relation. A cat-picture "is something for which it is intended that for those seeing it in its intended viewing conditions, it is F-related to the concept of being" a cat.[22] The signal feature of this account is reference to perceptual similarity for an agent. This can be taken as primitive, but it is plausible that it should be worked out in terms of identity or similarity of perceptual-cognitive states, perhaps neural states. Strikingly, in the definition of depiction, there is no reference at all to the kind of geometric projective semantics that was discussed in Section 2 and Section 5. The place of the geometry of projection is taken by the effect of a scene on a viewing agent. To this is added the hypothesis that pictures have a similar perceptual effect on viewers as the scenes they depict.

It is interesting to contrast Peacocke's characterization of depiction with Ross's. In applying Peacocke's approach to the semantics of the verb *depict* in (56), one compares the effect of p on viewers to the effect on viewers of actual cats on actual couches. As explained in Section 5, Ross proceeds by assigning a semantic value to p, and then checking whether that entails a proposition contributed by the object—here presumably the proposition that there is a cat on a couch. These are very different approaches.

(56) Picture p depicts a cat on a couch.

In this comparison, it is worth noting that Peacocke's and Ross's accounts are competing formalizations of a de dicto reading of (56). In addition, sentences describing the content of pictures have de re readings. Zimmermann (2006) analyzed four readings of (57), two of which he explains like this. "Whenever Pitchstone painted a bridge, he watched it from a convenient distance and applied paint of various colours to the surface of a canvas, which by the end of the day he had turned into a picture of the bridge (though, as we will see in a second, not always an entirely accurate one); in other words, Pitchstone portrayed bridges. Whenever Edlon painted a bridge, he did something similar, except that there was no bridge before his eyes, and if there was one before his mental field of vision, it did not derive from memory; in other words, there was no bridge that Edlon painted

26

a picture of, though he produced quite a few bridge-paintings." The Pitchstone reading is a de re reading, and the Edlon reading is a de dicto reading, which does not agree in detail with either Ross or Peacocke. For Peacocke, we could say that whenever Pfau painted a bridge, he applied various colors to the surface of a canvas, intending that the artifact he thereby created would for typical viewers be F-related to the concept *bridge*.

(57) He painted a bridge.

One area where the perceptual-cognitive approach may have an advantages is with non-realistic pictures such as caricatures. A purely geometric approach seems to have trouble in explaining the connection between the real Donald Trump and a caricature of Trump like (58). Gombrich (1977) discussed this problem and hypothesized that in perceptual-cognitive terms, a caricature of Trump *is* similar to Trump.

(58)

Caricature by DonkeyHotey. Creative Commons attribution license. Source images are creative commons licensed from Gage Skidmore's flickr photostream and smilygrl's flickr photostream.

Returning to transparency, it is clear that in our world and experience, it applies literally only in restricted situations such as architectural trompe l'œil features and constructions with peepholes. More commonly, when viewing a picture, our visual experience is not the same as when viewing one of the scenes compatible with the semantic content of picture, or the scene which is assumed to be depicted. We can see that we are viewing a picture and not one of the depicted scenes. There are many reasons for this. Color paintings and color photos that are viewed in reflected light result in a narrower range of intensity than what results from viewing a natural scene. Our eyes are focused only within a limited depth of field when viewing a natural scene, but can be focused over all of the surface of a picture. A picture must be viewed from a particular point to obtain an approximation of the scene that is assumed to be depicted. Viewing a black and white photo of a fire engine or a line drawing of it is not perceptually equivalent to viewing a red fire engine. And so on. These facts are not necessarily a problem for perceptual approaches, because two perceptual states can be similar in some respects, without being identical.

Going a different direction than perceptual approaches, Kulvicki (2006) suggested a "reflexive" version of transparency which takes the perceiving agent out

of the definition. He starts with the observation that by choosing the right vantage point and camera parameters, one can take a picture of a picture and obtain a result that is identical to the original picture. He goes on to define a notion of transparency that refers to the possibility of the object of a representation (such as a picture) being isomorphic to the representation. Kulvicki does not make the semantic assumptions stated in Section 2. Nevertheless in that framework one can say that a representational mode is transparent if and only if for any representation p in the mode, p is a part of some described situation for p. In the case of pictures, the picture p is a part of some situation in the set of situations $[\![p]\!]$. This notion of transparency has the virtue or feature of eliminating reference to perceivers, so that we do not have to talk about perceptual psychology to talk about transparency.

8. Discussion

This chapter has worked through a possible worlds semantics for pictures, and looked at applications to temporal progression, indexing, and other problems. Some differences between this enterprise and possible worlds semantics for sentences have emerged. The projective method for obtaining propositional semantic values is different from what is seen in the semantics of language, in that it is not syntax-driven, and involves no substantial composition of semantic values. Temporal progression and indexing seem to be more pragmatic in pictorial narratives than in linguistic ones. Nevertheless, the end products of semantic and pragmatic interpretation are similar in the two cases. This suggests that pragmatic, discourse-structural, or "post-semantic" solutions may be more viable for linguistic problems than linguistic semanticists often assume.

A viewpoint-centered semantics for pictures is supported by the linguistic argument reviewed in Section 5, and independently by the account of indexing in Section 4 and the consideration of ambiguity of pictures in Section 6. Viewpoint-centering in the semantics of pictures, in linguistic descriptions of pictures, and in other viewpoint-sensitive linguistic constructions are an intriguing connection between the semantics of pictures and the semantics of sentences.

In discussing possible worlds semantics for pictures, the intent is not to implicitly claim that this approach is superior to other ways of theorizing about the semantics of pictures, such as the perceptual-cognitive approach mentioned in Section 7. It seems though that possible worlds semantics for pictures on a geometric basis is an interesting meeting point for issues and methods emerging from art history, comics studies, computer science, linguistics, philosophy, and psychology.

Notes

[1]Thanks to Mats Rooth and Ede Zimmermann for comments and discussion, and to Milka Green for assistance with artwork. I am grateful to the artists who extended permission for reproduction of their work, and to their representatives. Ramona Hiller did invaluable editorial work in coordinating these permissions.

[2]See Article 136 ('Representing Intensionality').

[3]In the semantics for language, the object inside the brackets is usually taken to be not a simple sequence of words but a syntactic tree, or other representation of the syntax of a sentence. Section 6 brings up the possibility that the objects inside the semantic brackets should have extra structure also for pictures.

[4] This is the strategy of proxy possible worlds, where mathematical structures are used as possible worlds. Sider (2002) discusses metaphysical views about possible worlds and how these interact with the application of possible worlds in semantics.

[5]The geometric pictures in this section were produced in Matlab using the Geom3d package (Legland 2009). From a certain viewpoint, both w_2 and w_3 render to the same picture p_2. From a viewpoint with a higher vertical coordinate, w_3 renders to p_3, but w_2 does not, because w_2 has no octahedron.

[6] Section 6 takes up the possibility that instead of the viewpoint being existentially quantified, the semantic value of a picture is a set of centered worlds, consisting of a pair of a world and a viewpoint.

[7]Or in the semantics from Section 5, each picture contributes a set of centered situations.

[8]See Article 33 ('Aspectual Composition') and Article 38 ('Lexical Aspect').

[9] See Kamp and Rohrer (1983), Hinrichs (1986), and Kamp and Reyle (1993) for specific proposals.

[10]This conclusion applies to pictorial semantic content as formalized in Section 2. But classic comics and manga often include panels with "motion lines" that correspond to the velocity of an object. Such panels presumably have a non-stative literal content. A separate point about (20) is the possibility of a special "free perception" or "point of view" interpretation, where the bobcat picks up information corresponding to the second picture. On this, see Abusch and Rooth (2017).

[11]Such a theory was developed in Heim (1982).

[12]This example was rendered in Blender from data-structural scene descriptions. Artist Mats Rooth.

[13] Ross assumed that the pictures do have the same propositional content. Rooth and Abusch (2017) argue that this is not correct in projective semantics as reviewed in Section 2, but that the argument is saved by certain assumptions about marking rules and the space of worlds.

[14]See Section 5 of Article 8 ('Attitude Verbs').

[15] While de se readings can be illustrated also for pronouns in attitude descriptions, examples with PRO are especially convincing, because they have unambiguous de se readings (Chierchia 1989).

[16] See Aurnague et. al. (2001).

[17] On the opposite conclusion, see Ross (1997:Section 5.4). In that section, Ross discusses an account where a viewpoint is not merely geometric, but amounts to a visual experience.

[18] Reference to trees is a placeholder for whatever counts as a syntactic structure in a given syntactic theory. In addition to a conventional syntactic tree, this can be for instance a graph-shaped feature structure (Pollard and Sag 1994), or a minimalist derivation (Stabler 1997), or a categorial-grammar derivation (Steedman 2000).

[19]Or if time is included, a set of triples of worlds, times, and viewpoints.

[20]This actually requires some work, because as described they are features on line segments, not on points.

[21] Since there are many scenes that are consistent with the semantic content of a picture, it is wrong to refer to the perceptual experience of viewing a unique depicted scene. Instead we can say that the perceptual experience of viewing the picture is equivalent to the experience of viewing any of the scenes that are consistent with the semantic content of the picture.

[22]Peacocke (1987), page 388.

References

Abusch, Dorit. 2005. Causatives and mixed aspectual type. In Greg Carlson and Jeffrey Pelletier (eds.), *Reference and Quantification: the Partee Effect*. CSLI Publications.

Abusch, Dorit. 2012. Applying discourse semantics and pragmatics to co-reference in picture sequences. Sinn und Bedeutung 17.

Abusch, Dorit. 2014. Temporal succession and aspectual type in visual narrative. In Luka Crnč and Uli Sauerland (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, vol. 1, MITWPL 70, 2014.

Abusch, Dorit and Mats Rooth. 2017. The formal semantics of free perception in pictorial narratives. Proceedings of the 21st Amsterdam Colloquium.

Aurnague, Michel, Myriam Bras, Laure Vieu, and Nicholas Asher. 2001. The syntax and bsemantics of locating adverbials.

Beck, Sigrid and Arnim von Stechow. 2014. Events, Times and Worlds—An LF Architecture. Manuscript, Tübingen.

Bittner, Maria and Ken Hale. 1995. Remarks on definiteness in Warlpiri. In E. Bach, E. Jelinek, A. Kratzer, and B. Partee (eds.), *Quantification in Natural Languages*, 81-105.

Blender Institute. 2010. *Sintel*. URL `youtube.com/watch?v=eRsGyueVLvQ`. Accessed Jan. 1, 2018.

Blumson, Benjamin 2010. Pictures, perspective, and possibility. Philosophical Studies 149: 135-151.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner and Martin Walkow (eds), Proceedings of NELS 37, pp 101-114. Amherst: University of Massachusetts, Graduate Linguistic Student Association.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Blackwell.

Cheng, Lisa Lai-Shen and Sybesma, Rint. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. Linguistic Inquiry 30.4.

Chierchia, Gennaro. 1989. Anaphora and attitudes de se. In Renate Bartsch, J. F. A. K. van Benthem, P. van Emde Boas (eds.), *Semantics and Contextual Expressions*.

Chomsky, Noam 1980. On Binding. Linguistic Inquiry 11.1, pp. 1-46.

Clowes, Max. 1971. On seeing things. *Artificial Intelligence 2* pp. 79-116.

Cohn, Neil. 2013. Navigating comics: an empirical and theoretical approach to strategies of reading comic page layouts. Frontiers in Psychology 4:186.

Dekker, Paul. 2012. *Dynamic Semantics*. Springer.

Despic, Miloje. 2013 Binding and the Structure of NP in Serbo-Croatian. Linguistic Inquiry 44.2.

Dowty, David. 1979. *Word Meaning and Montague Grammar*.

Gombrich, Ernst. 1977. *Art and Illusion: A Study in the Psychology of Pictorial Representation*.

Greenberg, Gabriel. 2011. *The Semiotic Spectrum*. PhD dissertation, Dept. of Philosophy, Rutgers University.

Greenberg, Gabriel. 2013. Pictorial Semantics. Philosophical Review 122:2.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation, UMass Amherst.

Hinrichs, Erhard. 1986. Temporal anaphora in discourses of English. Linguistics and Philosophy.

Huffman, David. 1971. Impossible objects as nonsense sentences. Machine Intelligence 6, pp. 295-323.

Kamp, Hans. 1981. A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, M. Stockhof (eds.) Formal Methods in the Study of Language. Amsterdam: Mathematics Center, pp. 277-322.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Kamp, Hans and Christian Rohrer. 1983. Tense in texts. In Rainer Bäuerle, Christoph Schwarze and Arnim von Stechow (eds.), *Meaning, Use and Interpretation of Language*. De Gruyter.

Karttunen, Lauri. 1969. Discourse referents. Coling, pp. 1-38.

Krifka, Manfred. 1989. *Nominalreferenz und Zeitkonstitution*. W. Fink.

Kulvicki, John. 2006. *On Images: Their Structure and Content*. Oxford University Press.

Legland, David. 2009. Geom3d. URL `matgeom.sourceforge.net`. Accessed Jan. 1, 2018.

Lewis, David. 1979. Attitudes de dicto and de se. Philosophical Review.

McCloud, Scott. 1994. *Understanding Comics: the Invisible Art*. HarperPerennial.

Necker, Louis. 1832. Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. London and Edinburgh Philosophical Magazine and Journal of Science 1.5: 329337.

Peacocke, Christopher. 1987. Depiction. Philosophical Review 96.3, pp. 383-410.

Perry, John. 1977. Frege on demonstratives. Philosophical Review 86 pp. 474-97.

Pollard, Carl and Ivan Sag. 1994. *Head Driven Phrase Structure Grammar*. University of Chicago Press.

Roman, Augusto Gaurav Garg and Marc Levoy. 2004. Interactive Design of Multi-Perspective Images for Visualizing Urban Landscapes. IEEE Visualization.

Rooth, Mats and Dorit Abusch. 2017. Picture descriptions and centered content. Proceedings of Sinn und Bedeutung 21, 2017.

Ross, Jeff. 1997. *The Semantics of Media*. Kluwer.

Sedgwick H.A. 1980. The geometry of spatial layout in pictorial presentation. In Margaret Hagen (ed.) *The Perception of Pictures*. New York: Academic Press.

Sider, Theodore 2002. The ersatz pluriverse. Journal of Philosophy 99: 279-315.

Stabler, Edward. 1997. Derivational minimalism. In C. Retor, ed. *Logical Aspects of Computational Linguistics*, pp. 68-95. Springer.

Steedman, Mark. 2000. *The Syntactic Process*. MIT Press.

Strothotte, Thomas and Stefan Schlechtweg. 2002. *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann.

Szeliski, Richard. 2010. *Computer Vision: Algorithms and Applications*. London: Springer.

Tanaka, Masahi. 1992. *Gon*. Kodansha Limited. Current edition Kodansha Comics, 2011.

Tanaka, Takamasa, Kenji Shoji, Fubito Toyama and Juichi Miyamichi. 2007. Layout analysis of tree-structured scene frames in comic images. Proceedings of IJCAI.

Tatsumi, Yoshihiro. 2008. *Gekiga Hyōryū* (A Drifting Life). English edition Drawn and Quarterly, 2009.

Taylor, Barry. 1977. Tense and continuity. Linguistics and Philosophy 1, 199-220.

Tezuka, Osamu. 1970-1971. *Kirihito Sanka* (Ode to Kirihito) Part 2. English edition Vertical, 2006. Originally serialized in Big Comic.

Waltz, David Leigh. 1972. *Generating Semantic Description from Drawings of Scenes with Shadows*. PhD thesis, MIT.

Zimmermann, Ede. 2006. Quant paint. In Hans-Martin Gärtner et al. (eds.), *Between 40 and 60 Puzzles for Krifka*.

Zwicky, Arnold and Jerrold Sadock (1975). Ambiguity tests and how to fail them. In Kimball (ed.) *Syntax and Semantics 4*. New York:Academic Press.