

Penultimate proof version. (Page numbers of published chapter in brackets in grey font.)

Published as Chapter 4 in Bradley Armour-Garb, ed. *Reflections on the Liar*.  
Oxford: Oxford University Press, 2017. DOI:10.1093/oso/9780199896042.003.0004

# **Gestalt Shifts in the Liar**

## **Or**

# **Why KT4M Is the Logic of Semantic Modalities**

Susanne Bobzien

**ABSTRACT.** This chapter offers a revenge-free solution to the liar paradox and presents a formal representation of truth in, or for, a natural language like English, which proposes to show both why (and how) truth is coherent and how it appears to be incoherent, while preserving classical logic and most principles that some philosophers have taken to be central to the concept of truth and our use of that notion. The chapter argues that, by using a truth operator rather than truth predicate, it is possible to provide a coherent, model-theoretic representation of truth with various desirable features. After investigating what features of liar sentences are responsible for their paradoxicality, the chapter identifies the logic as the normal modal logic KT4M (= S4M). Drawing on the structure of KT4M (=S4M), the author proposes that, pace *deflationism*, truth has content, that the content of truth is bivalence, and that the notions of both truth and bivalence are semideterminable.

**KEYWORDS.** truth, liar paradox, modal logic, S4M, duck-rabbit illusion, Gestalt shift

The guiding question for this chapter is what modal system with a sentential truth-operator—if any—would capture the notion of truth in a semantically closed natural language. Or, more

modestly, what modal system with a sentential truth-operator—if any—would capture the notion of truth in the propositional fragment of a semantically closed natural language? As a starting point toward an answer, I consider the basic features of liar sentences that, combined, create the liar paradox and that threaten the consistency of the notion of truth. Through distilling the structure from these features, and separating pragmatic from nonpragmatic elements, I obtain the building blocks for a possible-world semantics that leads to the normal modal logic KT4M (or S4M or S4.1). This logic, which so far has gained little attention, makes it possible to represent truth in the propositional fragment of natural language as a coherent notion. It also provides some additional insight into the notion of truth, such as that, *pace* deflationism, truth does have content and that this content is bivalence; moreover that it appears that the sets of true sentences and of bivalent sentences are semi-determinable on the propositional fragment of natural language. Accordingly, I offer KT4M as the logic of semantic modalities, with truth as semantic necessity and bivalence as semantic noncontingency. What I suggest is *not* an axiomatic theory of truth.

The route from the Liar to KT4M includes the following way-posts. The paradoxical liar sentences lie at the intersection of three sentential features: First, their ascription of a semantic value, second, their self-reference, and third, their predicating something that is incompatible with truth. Individually and in pairs any of these features are unproblematic, and it is an indication of the richness of natural languages that they include all three. Things go awry when the three features are exemplified in the same sentence (or in a plurality of sentences related by anaphora or a successor function).

Sentences (72) that self-ascribe a semantic value, such as the liar sentence<sup>1</sup>

(L) This sentence is false

and the truth-teller sentence

---

<sup>1</sup> Why do I not use ( $L_{\text{untrue}}$ ) ‘This sentence is untrue’ as the liar sentence? The answer will become clear later. ( $L_{\text{untrue}}$ ) is discussed in 4.11. Why do I not use a sentence like (L101) ‘The sentence written on the board in room 101 expresses a false statement’ as an example of a liar sentence? The answer is brevity. Throughout, you should be able to replace (L) by (L101), and (T) by a corresponding (T101) without any philosophically significant change in what I say.

(T) This sentence is true

are conceptually bistable on account of salience (section 4.2). Sentences that self-ascribe a semantic value also display a context sensitivity that is based on the fact that the designation of their subject expression is an object that itself has truth conditions. Such context sensitivity is displayed both by (L) and by (T) (section 4.4). Sentences that self-ascribe a semantic value that is incompatible with truth, such as (L), display in addition to context sensitivity a kind of unsavoury assessment sensitivity (section 4.5). This assessment sensitivity has as a consequence that the appropriate epistemic position toward liar sentences is an—iterating—suspension of judgment concerning their semantic value (section 4.6). The combination of the structural features of salience-based bistability, context sensitivity, and assessment sensitivity of liar sentences leads to the modal system KT4M as the correct choice for modeling truth as a coherent feature of natural language. It provides informal correlates to the axioms of KT4M as well as the basic elements for an interpretation of the corresponding Kripke semantics (section 4.7-4.9). The paradoxicality of the Liar finds an explanation in the confusion of an elusive pragmatic element with a semantic or logical one in what is known as the liar property, and it is dissolved in a revenge-proof manner (section 4.10).

The chapter is something of a pioneer piece, with rough edges and uncharted trails. It invites the reader to explore thinking about truth in (yet) a(nother) new way. This notwithstanding, it draws upon numerous aspects of established approaches to the Liar, and displays similarities, some of them strong, to various well-known theories of truth. It provides functional analogues for Kripke's fixed points and builds on Herzberger's and Kripke's notion of ungroundedness by giving it a contextualist explanation. It shares with more recent axiomatic theories and revisionist theories of truth their desire to keep logic classical, semantics bivalent, and truth untyped, as well as their use of modal(-like) axioms to represent (some of) the structure of truth. Its closest cousin among well-established axiomatic theories of truth is Kripke-Feferman, and its closest revisionist cousin, and possibly closest relative altogether, is Herzberger. Thus, the identification of bivalence as the content of truth may already be implicit in Herzberger 1982a. Moreover, this chapter offers as alternative to Herzberger's *semantic* revision steps for the Liar a *pragmatic* oscillation whose significance is bound up with its occurrence in arguments. The pragmatic oscillation can be linked to Barwise

and (73) Etchemendy’s ‘simply false’ and ‘simply true’, but instead of a blunt ambiguity, I suggest a shift in conceptual salience. The chapter adopts McGee’s approach of treating truth as a sentential operator, but without giving rise to a revenge problem—or so I hope. Like van Fraassen’s supervaluationist fixed-point models of truth, it separates bivalence and the law of excluded middle, but unlike his theory, it retains compositionality and does not entail a third semantic value. The chapter also picks up on—and turns around—recent suggestions that provide a possible-world semantics for truth and treat necessity and truth in similar ways (Leitgeb 2003; Halbach 2003; Halbach, Leitgeb, and Welch 2005) and shares with Billon 2011 the treatment of arguments as contexts. Beyond the amalgamation of all these factors into one theory, there are the features mentioned in the first two paragraphs.

Many questions that arise for theories of truth generally are not, or hardly, touched upon in this chapter: these include related semantic paradoxes such as Curry’s and Yablo’s, the Gödelized Liar (which gets its own paper!), and the relations to quantified truth-ascriptions and propositional attitude ascriptions. This is neither to say that answers to these questions cannot be given nor that they will not be given. Finally, the chapter is deliberately written in such a way that it is accessible to nonmathematicians and it keeps formal elements to a minimum.

#### 4.1. Truth Predicate and Truth Operator

In natural language, there are two main uses of the expression ‘true’ with which truth is ascribed to a suitable truth-bearer: a predicative use, as in ‘that’s true’, ‘the Barcan Formula is true’, ‘whatever she says is true’, and a sentential-operator use, as in ‘it is true that it is raining’, ‘it is true that he is both charming and annoying’. Truth as a predicate commonly requires either a noun like ‘sentence’, ‘proposition’, etc., or a name or description for an instance of a sentence, proposition, etc., or is anaphoric. Truth as a sentential operator generally neither requires nor allows these. There is no straightforward translation mechanism from one natural-language use to the other. Biconditionals such as ‘the sentence ‘ $\ulcorner S \urcorner$ ’ is true (where ‘ $\ulcorner S \urcorner$ ’ is the name of the sentence abbreviated as  $S$ )  $\leftrightarrow$  It is true that  $S$ ’ provide no more than a rough guide for how to move between the two kinds of truth ascription.<sup>2</sup> Predicatively

---

<sup>2</sup> I agree with Halbach 2003, p. 79, that “there is hardly any essential *syntactical* difference in English between ‘true’ and ‘necessary’, that is, replacing a used occurrence of ‘true’ by ‘necessary’ in an English sentence usually

truth-ascribing sentences with noun phrases such as ‘That he is both charming and annoying is true’ can be translated into truth-operator sentences and vice versa. For present purposes, any difference between these two is treated as semantically irrelevant.

In the proposed formalization of truth, truth is treated not as a predicate but as a sentential operator. The main reason is that this brings out some significant (74) structural features of truth in a straightforward manner. It makes it possible to put forward, for a language that is semantically closed in that it contains its own truth predicates or truth operators, the following: a coherent, model-theoretic representation of truth that preserves classical logic, does not go beyond normal modal logic, and in which liar sentences are consistent.

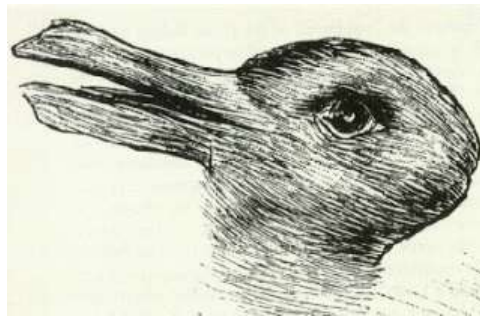
To be very clear: my treatment of truth ascription with a truth operator for truth in natural language is not meant to suggest that the natural-language use of a truth operator is in any way superior or ‘closer to the truth’ than that of a truth predicate. The structure of truth which I aim to define modally is meant to *capture equally* the operator use and the predicate use in natural-language discourse. Counter to axiomatic theories of truth (which favour a truth-predicate), I suggest that the structure of truth itself disqualifies there being a semantic Liar property, and the way it does so is best expressed with a truth operator. In my view, the reasons why speakers use one or the other are generally pragmatic. Linguistic exploration of the contrasting natural-language uses of ‘it is true that . . .’, ‘. . . is true’, ‘truth’, ‘. . . is false’, etc., may prove enlightening in various ways, but I believe that their usefulness to a solution of the semantic paradoxes is restricted.

## 4.2. Perceptual Multistability

---

will yield again a sentence and *vice versa*.” Where Halbach 2003 and Halbach, Leitgeb, and Welch 2005 suggest treating necessity as a predicate, I suggest treating truth as an operator that attaches to what is said in a sentence. I give a rough idea how this works. The sentence ‘the sentence “snow is white” is true’ is short for ‘the sentence “snow is white” says that snow is white and it is true that snow is white.’ The sentence “Fermat’s last Theorem is true” is short for ‘Fermat’s Last Theorem says that no three positive integers . . . and it is true that no three positive integers . . .’ The sentence ‘Everything the pope says is true’ is short for ‘for all sayable things  $p$  (where  $p$  is a sentential variable), if the pope says (that)  $p$ , then it is true that  $p$ .’ Or alternatively conjunctively ‘If the pope says that grass is green, it is true that grass is green and if the pope says that snow is white, it is true that snow is white, . . .’ More complex sentences and propositional attitude ascriptions are dealt with in the same general manner.

The topic of the next two sections is multistability. In this section I suggest that sentences that self-ascribe a semantic value are conceptually multistable. I explain the notion of conceptual multistability by way of the notion of perceptual multistability. That is, I use the analogy to perceptual multistability heuristically, to aid the reader in getting an understanding of what conceptual multistability is. (The following is not an argument by analogy.)



The (75) duck-rabbit (or rather duck-hare) illusion is a bistable perceptual phenomenon. It is an instance of a figure, or more generally a visual pattern, that permits a Gestalt shift. You look at it. You see a duck, no doubt about it. You look at it some more. Suddenly, instead you see a rabbit—then possibly again a duck. But you never see both at a time. Perceptual phenomena like the duck-rabbit illusion are called multistable, since they allow a perceiver to experience in succession two or more different and incompatible stable percepts or Gestalts, resulting from a so-called perceptual reversal or Gestalt shift. The incompatibility is on the side of the perceiver: the figure itself may depict both a duck and a rabbit, but the perceiver cannot simultaneously see (the depiction of) a duck and a rabbit. (The incompatibility seems to be a brute fact of perception. The details are irrelevant here.) Other examples of perceptual multistability are the Necker cube, the stacked cubes, the mother-father-daughter figure and the spinning dancer—a kinetic multistable figure.<sup>3</sup>

Here are 10 characteristics that all multistable perceptual illusions seem to share.

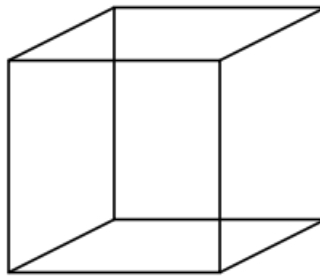
- (a) Which Gestalt one sees, or sees first, may depend on multiple factors, including simultaneous or preceding circumstantial factors.
- (b) In ordinary circumstances, whether a person sees the one Gestalt (a duck), or the other (a rabbit), or experiences a Gestalt shift from one to another, the person *cannot be faulted* for what they (say they) see. If someone is presented with a range of sketched animal representations on cards, with a duck-rabbit

---

<sup>3</sup> [http://en.wikipedia.org/wiki/Spinning\\_Dancer](http://en.wikipedia.org/wiki/Spinning_Dancer).

card snuck in, and is asked to say for each card what animal it shows and then move on to the next card, we have no reason to say that someone who says ‘a rabbit’ or someone who says ‘a duck’ when faced with the duck-rabbit card is mistaken. Someone saying ‘a rabbit when looked at one way, a duck looked at another way’ would not be mistaken either.

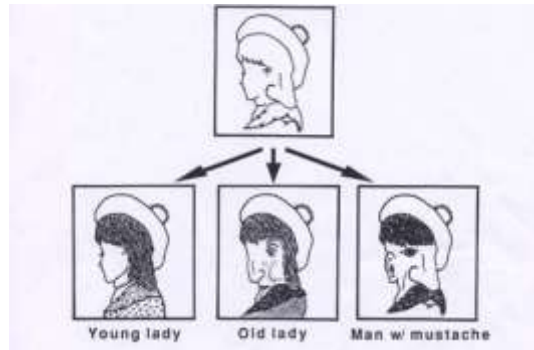
- (c) Likewise, a person cannot be faulted if, seeing one or the other Gestalt, they make some *inferential observations*: when you see the duck, you can make the inferential observation that part of the figure or drawing you see depicts a beak.
- (d) Some people only see one Gestalt and experience no Gestalt shift. In such cases, one can sometimes *prompt* a Gestalt shift in the person’s perception, e.g., by pointing out some features of the alternative Gestalt. One may be able to make someone who only sees the rabbit see the duck by saying ‘don’t you see the beak?’
- (e) Such prompts do not by themselves warrant that the perceiver ‘catches on’ and sees the alternative Gestalt. Multistable perceptual illusions can be stubborn, and are so to different degrees. With some illusions, for some people it remains impossible, or takes very long, to reverse the Gestalt, even if they are told what the second stable Gestalt is (or depicts). The Necker cube is a good example:



Even better is the *spinning dancer*. (76)

- (f) As a rule, one or more elements in the multistable figure do *double duty*: one and the same part of the figure may depict one thing relative to one stable percept, another relative to the other—e.g., a beak and a pair of ears.
- (g) In multistable perceptual phenomena each of their stable percepts is in some respect *deficient*—for example, since something has been left out (color, shading or missing details, etc.); or due to the two-dimensionality of the representation of something three-dimensional.

- (h) Perceptual multistability can be *removed*, e.g., by adding visual elements. In the case of the duck-rabbit illusion some feather details could be added. Here is an example with the mother-father-daughter figure that illustrates possibilities of the removal of multistability by adding detail:



- (i) Someone's experience of a Gestalt shift is in several respects independent of their having a theoretical grasp of it. It is perfectly possible to experience a Gestalt shift *either* without having a concept of multistable perception; *or* without having concepts of the Gestalts depicted (one may notice no more than a bird-to-mammal shift, or one-animal-to-another shift); *or* one may even be unaware that one experienced a shift of Gestalt. (The spinning dancer may at some point be perceived spinning counterclockwise instead of clockwise without the observer ever contemplating which way the figure spins, and consequently missing that it 'changed direction'.) (77)
- (j) As a result of (i), it may be difficult for us to ascertain which of two Gestalts in a bistable figure someone perceives; they may lack the ability to describe, or even fully understand, what it is they perceive. Someone may experience a Gestalt shift with the stacked cubes illusion, without ever having considered that they are looking at a figure that depicts two different (in arrangement and number) stacks of cubes, or any cubes at all.

### 4.3. The Multistability of Sentences That Self-Ascribe a Semantic Value



In parallel with the notion of perceptual multistability one can institute a notion of *conceptual multistability*.<sup>4</sup> Simple structural ambiguities make straightforward cases of conceptual multistability. Take (TEL) ‘I saw the woman with the telescope’. Without a context given, you may ‘read’ the sentence in one way (i) (the speaker had the telescope) or another (ii) (the woman had the telescope). Or (iii) you may shift from one reading to the other. The incompatibility is on the side of the ‘reader’: the sentence itself can be used to express either reading, but the reader cannot simultaneously read it both ways. (This potential for incompatible readings seems to be a brute fact of linguistic representation of natural languages.) It is left to the interested reader to verify that one can observe 10 characteristics that match those of perceptual multistability.

Conceptual multistability is not restricted to structural (and lexical) ambiguity. Sentences that predicatively self-ascribe a semantic value are also conceptually multistable, if less obviously so. Instead of a shift between reading one of two or more possible contents of a sentence, there is a shift between focus on one of two possible *aspects* of a sentence, with no change of content, or, more precisely, a shift between what is *salient* in that sentence. (To indicate that this multistability is a matter of salience of content, not content, I replace the expression ‘to read’ and its cognates that I used for ambiguities with ‘to understand’ and *its* cognates.) Salience is a pragmatic element of a more elusive kind than ambiguity.<sup>5</sup> Still, as an aid to grasping this salience-based notion of Gestalt shifts, one can think of Barwise and Etchemendy’s ambiguity-based situation shifts, or of the revision steps in Herzberger’s revision semantics.

Semantic-value-self-ascribing atomic sentences (i) can be understood designationally as that which is denoted by the designator of a semantic-value-self-ascribing sentence, where this denotation is *left unanalyzed* (beyond its being syntactically composed of a predicate and a designator); or (ii) they can be understood as sentences that *ascribe a truth value*; again, (iii)

---

<sup>4</sup> More accurately, this would be the potential for shifts in the perceptual salience of linguistic stimuli.

<sup>5</sup> In this chapter I do not further specify this pragmatic element of attention or focus and perceptual salience of linguistic stimuli. The details are in part empirical. (For an overview of some recent research see Summerfield and Egner 2009.) Let it suffice to say that what I have in mind is neither a difference in force nor one in presuppositions nor one in implicatures.

it is possible to shift from one understanding to the other. Liar and truth-teller sentences are bistable sentences of this sort.

The (78) alternative Gestalts of such bistable sentences are less easily detected than in the case of the ambiguity of (TEL). They are best made apparent contextually; that is, by means of typical contexts in which each will—almost certainly—occur. The most effective linguistic contexts are arguments. For understanding (i), consider the following derivation:

1. (L) is true. assumption
2. (L) is false. 1, semantic descent

Here, in order to see that and how 2 is derived in this derivation, one needs to think of 2—at least temporarily—as that which is denoted by the designator in 1. Or take

1. (L) is false. assumption
2. (L) is true. 1, semantic ascent

Here, in order to see that and how 2 is derived in this derivation, one needs to think of, and thus make use of, 1 as that which is denoted by the designator in 2. For the ascriptional option, (ii), consider the following derivation:

1. (L) is false. assumption
2. (L) is true. 1, semantic ascent
3. Contradiction<sup>7</sup> 1, 2, principle of bivalence

This is part of an informal version of the Liar. If you succeed in seeing the—presumed—contradiction, then at least while moving to line 3, you have understood (L) in line 1, as ascriptional. You have made use of the fact that (L) self-ascribes a truth value. The liar paradox also (iii) provides us with an *example of a conceptual Gestalt shift* within an argument. Consider this informal version of the simple liar paradox (SLP):

---

<sup>6</sup> There is no *formal* contradiction here. Rather, it results from the acceptance of the Principle of Bivalence. If instead of (L) one uses (L<sub>untrue</sub>) ‘This sentence is untrue’, one gets a formal contradiction. A Gestalt shift is still required.

1.	(L) is true.	assumption 1
2.	(L) is false.	1, semantic descent
3.	If (L) is true, (L) is false.	1, 2, discharge assumption 1
4.	(L) is false.	assumption 2
5.	(L) is true.	4, semantic ascent
6.	If (L) is false, (L) is true.	3, 4, discharge assumption 2
7.	(L) is true if and only if (L) is false.	3, 6, definition ‘if and only if’
8.	(L) is true or false, and not both.	principle of bivalence
9.	Contradiction	7, 8

Someone (79) who grasps (SLP) needs to understand ‘(L) is false’ (i) as the denotation of the designator of a semantic-value-ascribing sentence for the steps from 1 to 2 and from 4 to 5. They need to understand ‘(L) is false’ (ii) as ascribing a truth value for the step from 7 and 8 to 9. A person who entertains (SLP) would thus experience at least one Gestalt shift somewhere between lines 1 and 9.

As in the case of perceptual bistability, the incompatibility of (i) and (ii) is on the side of the person who entertains the sentence, e.g., as part of an argument. Sentence (L) lends itself to either understanding, but one cannot simultaneously entertain (L) in both ways. (Again, this potential for incompatible understandings seems to be a brute fact regarding semantic-value-self-ascribing sentences.) Again, the incompatibility at issue seems psychological. It appears to be psychologically impossible for us to simultaneously entertain both understandings. The following deliberation may illustrate this. Take

1                    (L) is false.

Here (L) can be understood (i) as that which is denoted by ‘(L)’ in 1 or (ii) as 1. Now, if we entertained (i) and (ii) simultaneously, we would have to include the fact that what is denoted by ‘(L)’ in 1 is 1. Thus we would have instead of (1)

1’                    ‘(L) is false’ is false,

where (L) is understood simultaneously both (i) as what is denoted by ‘(L)’ and (ii) as 1’. Then we would have to include the fact that what is denoted by ‘(L)’ in 1’ is 1’. Thus we would have

1”            ““(L) is false’ is false’ is false’ is false.

And so forth. Arguably, it is psychologically impossible to entertain this infinite item.<sup>8</sup>

As with perceptual multistability, so with conceptual multistability, both the Gestalt understood, and the Gestalt shift, have to be experienced by the relevant individuals themselves. Evil neuroscientists aside, one cannot force someone to understand a sentence one way rather than the other, nor can one force them to experience a Gestalt shift. What one *can* do is provide trigger elements—as I just did. (It follows that I need to rely on your, the reader’s, cooperation.)

Moreover, the salience-based bistability of semantic-value-self-ascribing sentences displays all ten characteristic features of multistability. Again, with (L) as example:

- (a) Whether you understand (L) (i) as designational or (ii) as ascriptional when you encounter it is likely to depend on various circumstantial factors, such as steps in arguments (see above). (80)
- (b) We would neither fault someone who understands (i), nor someone who understands (ii), nor someone who states that (L) can be understood both as (i) and as (ii). This can be seen from our acceptance of the arguments that illustrate each of these cases.
- (c) Someone would not be faulted if they inferred ‘(L) is false’ from ‘(L) is true’, using (i). Someone would not be faulted either, if they inferred ‘(L) is not true’ from ‘(L) is false’, using (ii). Nor would someone be faulted for pointing out that depending on how they understood the sentence, they could infer different things.
- (d) We may encounter someone who, on their own, managed only to understand (i), or (ii), but can experience the alternative upon prompting.

---

<sup>7</sup> See Ryle (1951) for a related infinite iteration.

- (e) It can be difficult for someone to shift Gestalt, if they are simply told of the two options, without context, or if they are simply told that (L) is conceptually multistable.
- (f) The element of the linguistic representation that does *double duty* is the whole sentence. Relative to the designational understanding (i), it is what is denoted by the subject expression of the sentence. Relative to the ascriptional understanding (ii), it is the sentence.
- (g) What is the *deficiency* of (each of) the two understandings?<sup>9</sup> Usually, when we have a sentence, what is denoted by its subject expression and what truth value is ascribed to it come apart. Not so in semantic-value-self-ascribing sentences. They are—pragmatically—*deficient* in that they do not indicate which of these aspects is at issue.
- (h) We can *remove* the conceptual multistability of content *aspect* (or saliency) by adding explicitly whether the focus is on the sentence qua denotation of the subject expression or qua truth-value ascription.
- (i) In order to experience a Gestalt shift from (i) to (ii) or vice versa, someone would not need any knowledge *either* of content *aspects* (or saliency) *or* of predication or truth-value ascription; and *moreover* they may be unaware that they experienced a Gestalt shift. For instance, anyone who considers the liar paradox paradoxical experiences a conceptual Gestalt shift (see the simple liar paradox (SLP)).
- (j) As a result of (i), it may be difficult to ascertain in which Gestalt someone is understanding the sentence, as they may lack the ability to describe, or even comprehend, what their understanding is.

So, sentences like (L) are conceptually bistable. This kind of salience-based bistability is present whenever a sentence self-ascribes a truth value. It is a consequence of the combination of self-reference and truth-value ascription. Thus we can also have bistability in (T). Since supposition of (T) does not lead to contradiction, the bistability is harder to detect. Compare the following two derivations: (81)

---

<sup>8</sup> It is *not* the lack of semantic evaluability due to ungroundedness, for which see section 4.4.

- |    |              |                     |
|----|--------------|---------------------|
| 1. | (T) is true. | assumption          |
| 2. | (T) is true. | 1, semantic descent |
|    |              |                     |
| 1. | (T) is true. | assumption          |
| 2. | (T) is true. | 1, semantic ascent  |

The first derivation may evoke a descriptive understanding, the second an ascriptional understanding. Line 2 is the same in either case, but since it is derived in different ways, a reader would invoke different saliency for the understanding of either derivation. So (L) and (T) are both conceptually bistable.

A comparison of the distinction between ambiguity and salience-based bistability with that between the duck-rabbit on the one hand and the Necker cube and spinning dancer on the other is instructive. In the cases of the duck-rabbit illusion and ambiguity, we have a Gestalt shift between the representations of two different objects. In the cases of the Necker cube and spinning dancer illusions and semantic-value-self-ascribing sentences, we have a Gestalt shift not between two objects, but between two different perspectives on the same object (granting some convenient metaphysical assumptions about the identity of cubes). The relevance of this distinction becomes clear in section 4.10.

Overall, I do *not* suggest that the bistability of (L) per se explains its paradoxicality. It cannot. This is so because all sentences that predicatively self-ascribe a semantic value are bistable by saliency, and not all of them are paradoxical.<sup>10</sup> Rather, I argue that the bistability of liar sentences is *one of several factors* that are jointly sufficient to explain (L)'s paradoxicality and that have to be taken into account in a theory of truth.

#### **4.4. Self-Reference and Context Sensitivity in Liar and Truth-Teller Sentences**

---

<sup>9</sup> Of course, one may propose that we solve the Liar by prohibiting such salience-based Gestalt shifts. However, first, it seems entirely ad hoc, if one restricts such prohibition to liar sentences. Second, either one has to grant pragmatics an unusual impact on inferences, or one has to hold that there is a difference in linguistic content between ascriptional and designational understanding. Neither seems a good idea. (Some more on this in section 4.10.)

(This section is a compressed and criminally simplified presentation of a point I hope to set out in more detail in a separate paper.) A second feature that semantic-value-self-ascribing sentences share is what, following Herzberger 1970, Kripke has called their ungroundedness (Kripke 1975). My preference is to think of this feature as a kind of context sensitivity. Take the sentence

(2) The sentence ‘*s*’ is true.

This sentence can only be semantically evaluated if either (i) things are as the sentence denoted by the subject expression of (2) (i.e., the sentence ‘*s*’) says they are, or (ii) things are not as the sentence ‘*s*’ says they are. In the first case (2) is true, in the second it is false. I call the contexts in which (2) can be semantically (82) evaluated *circumstances*. They are (i) the circumstances in which things are as the sentence designated by the subject expression of (2) says they are, and (ii) the circumstances in which things are not as the sentence designated by the subject expression of (2) says they are. The assumption is that a circumstance is of the second kind precisely if it is not of the first kind, *tertium non datur*.

Accordingly, it is a necessary condition for the semantic evaluation of the semantic-value-self-ascribing sentences (T) or (L) that either (i) things are as the sentence denoted by their subject expression, i.e., the sentence (T) or (L), says they are, or (ii) things are not as the sentence denoted by their subject expression, i.e., the sentence (T) or (L), says they are. The contexts in which (T) or (L) can be evaluated are (i) the *circumstances* in which things are as the sentence designated by the subject expression of (T) or (L) says they are, and (ii) the *circumstances* in which things are not as the sentence designated by the subject expression of (T) or (L) says they are. I understand the assignment of a circumstance of one of the two kinds to the subject expression of sentence like (L), (T), etc., as the mapping of the subject expression onto a circumstance of one of the two kinds. This is in line with standard accounts of context sensitivity. Of course, the circumstances in which things (i) are, or (ii) are not, as the sentence designated by the subject expression in (L) or (T) says they are, are precisely the circumstances in which things (i) are, or (ii) are not, as the sentence *itself* says they are. This is another manifestation of the *self-reference* in (L), (T), etc. Therefore, if you prefer to think of such assignments of circumstances as—possibly random—stipulations regarding the sentences (L), (T), etc., *themselves*, rather than regarding their subject expressions, this is fine, too. The difference is immaterial for the formal representation of truth in sections 4.7 and 4.8. Either way, the assignment satisfies a necessary condition for the semantic evaluability of (L) or (T)

and thus, if you like, ‘grounds’ them in some sense. In function (not in kind) you can compare these assignments of a circumstance to (L) or (T) with the initial assignments of a semantic value to (L) or (T) in revision theories of truth (Herzberger 1982a and 1982b; Belnap and Gupta 1993).<sup>10</sup> Making semantic evaluability possible by fixing the context (or randomly stipulating how things are regarding (L) or (T)) is a required step that logically precedes the manifestation of assessment sensitivity in liar sentences to which I turn now.<sup>11</sup>

#### 4.5. Assessment Sensitivity

Liar and truth-teller sentences share the two features of predicative truth-value ascription and self-reference, which combine into truth-value self-ascription. Truth-value self-ascription both makes them bistable and gives them their specific context sensitivity. The third feature of liar sentences, the one that sets them apart from truth tellers, and—in conjunction with the shared features—is responsible for their paradoxicality, is that they ascribe a semantic value that is at odds with truth. In this section I describe how the confluence of all three features endows liar sentences with a specific kind of unsavory assessment sensitivity.

For this purpose, I introduce four very simple valid argument forms that produce conclusions (i) from the premise that things are as a sentence ‘*p*’ says they are and (ii) from the premise that things aren’t as the sentence ‘*p*’ says they are. There is a presupposition that the sentence ‘*p*’ says *something*. The name of the sentence ‘*p*’ is given in parentheses as (P):

(P)    *p*

I express the relation between sentence and name as ‘(P) says that *p*’. The four argument forms are then as follows:

##### *Argument form 1*

1.            Things are as (P) says.
2.            So (P) is true.

##### *Argument form 2*

1.            Things are as (P) says.

---

<sup>10</sup> Needless to say, such circumstances *are not* what Kripke calls ‘specifiable circumstances’ for his meaning criterion (Kripke 1975, p. 699).

<sup>11</sup> This kind of context sensitivity of (L) and (T) is quite unlike those suggested by Parsons (1974) and Glanzberg (2004).



2. So  $p$ .

*Argument form 3*

- 1 Things aren't as (P) says.
- 2 So (P) is false.

*Argument form 4*

- 1 Things aren't as (P) says.
- 2 So it's not the case that  $p$ .

*Argument forms 1 and 3* involve some kind of semantic ascent. *Argument forms 2 and 4* involve some kind of disquotation. (I use these argument forms, which I take to be generally accepted, in order to cast a shadow on the bivalence of the Liar sentence.)

Applied to (L), the four argument forms help to bring out the paradoxicality of (L) in contrast with the nonparadoxicality of (T). I construct four hypothetical arguments that move from the assumption of a fixed context of the liar sentence (section 4.4) to a semantic-value ascription to the liar sentence. In the first two arguments, the context for (L) is the circumstance that things are as (L) says (context  $c_1$ ), in the last two, the circumstance that things aren't as (L) says (context  $c_2$ ).<sup>12</sup>

**Argument 1**

- |    |                         |                              |
|----|-------------------------|------------------------------|
| 1. | Things are as (L) says. | context $c_1$                |
| 2. | So (L) is true.         | some kind of semantic ascent |

**Argument 2**

- |    |                         |                           |
|----|-------------------------|---------------------------|
| 1. | Things are as (L) says. | context $c_1$             |
| 2. | So (L) is false.        | some kind of disquotation |

**Argument 3**

- |    |                            |                              |
|----|----------------------------|------------------------------|
| 1. | Things aren't as (L) says. | context $c_2$                |
| 2. | So (L) is false.           | some kind of semantic ascent |

**Argument 4 (84)**

- |    |   |                           |
|----|---|---------------------------|
| 1. | Things aren't as (L) says.              | context $c_2$             |
| 2. | So it's not the case that (L) is false. | some kind of disquotation |
| 3. | So (L) is true.                         | bivalence                 |

---

<sup>12</sup> Since what (L) says is precisely what the sentence denoted by the subject expression of (L) says, for convenience I use the shorter formulation.

All four arguments appear valid (argument 4 on the explicit assumption of bivalence). I call them *legitimate*, by which I intend that they are of a form to which there are no counterexamples. If we assume that its first line is true, each argument is also per se irrefutable, that is, it cannot be directly refuted. Yet, for either context, there are two arguments with incompatible conclusions. Which conclusion one obtains depends on which argument one chooses. (If one produces a corresponding set of hypothetical arguments for (T), with either context, the arguments conclude with the same semantic-value ascriptions in either argument. No incompatible conclusions arise.)

Now, I call the fact that two different per se irrefutable arguments lead to conclusions that ascribe different semantic values to the same sentence ‘ $\ulcorner S \urcorner$ ’ *argument-based assessment sensitivity* of ‘ $\ulcorner S \urcorner$ ’. (L) is assessment sensitive in this sense. Moreover, if there are any two per se irrefutable arguments that ascribe in their conclusions two incompatible semantic values to the same sentence ‘ $\ulcorner S \urcorner$ ’, I assume that it is not true that ‘ $\ulcorner S \urcorner$ ’, or (in predicate formulation) that ‘ $\ulcorner S \urcorner$ ’ is not true *tout court*. I also assume that it is not false that ‘ $\ulcorner S \urcorner$ ’, or that ‘ $\ulcorner S \urcorner$ ’ is not false *tout court*. The reason is that neither the truth nor the falsehood of a sentence should depend on what per se irrefutable argument we use to infer that sentence’s truth value.

Let me present this argument-based assessment sensitivity of liar sentences somewhat more formally. Here is first, a general account of assessment sensitivity:<sup>13</sup>

- (3) An expression is *assessment sensitive* just in case its semantic value depends upon the viewpoint of assessment, i.e., the context from which it is assessed.

The only linguistic expressions at issue at this point are sentences (with the context of evaluation fixed as  $c_1$  or  $c_2$ ). And the semantic values at issue are not the truth and falsehood *tout court* of these sentences, but are values that function as *constitutive elements* of these—stand-alone—values. (This general idea should be familiar from supervaluationist theories.)

---

<sup>13</sup> Cf. e.g. MacFarlane (2014).

Since these values are constitutive of truth and falsehood *tout court*, I call them *semantic subvalues*. We then obtain this modified account.<sup>14</sup>

(4) A sentence (context fixed) is *argument-based assessment sensitive* just in case its semantic *subvalue* depends upon the argument at the viewpoint from which it is assessed.

Let (85) me explain the three main terms of (4) in terms of *argument forms 1 to 4*: I say that each *viewpoint* regarding an atomic sentence *S* houses a legitimate argument with a conclusion that is a semantic-value-ascription to *S*.<sup>15</sup> I call such atomic sentences *S* *input sentences* and such arguments *viewpoint-arguments*. These arguments are hypothetical deductions in a metalanguage.<sup>16</sup> Each viewpoint-argument for *S* terminates as soon as it reaches a semantic-value ascription to *S*. There are four relevant types of viewpoint arguments, and the argument forms 1 to 4 from above can serve to exemplify these, if we supplement them with additional uses of the principle of bivalence or semantic ascent where necessary. These arguments can be divided into two kinds, in line with section 4.2.3. First, there are *semantic-ascent-first viewpoint-arguments*. In these, a statement of the circumstances, (i) or (ii), is followed by the use of semantic ascent. For (L), *arguments 1* and *3* are such viewpoint arguments. Second, there are *disquotation-first viewpoint-arguments*. In these, a statement of the circumstances, (i) or (ii), is followed by the use of disquotation, and then by the use of further rules or theorems, if necessary. For (L), *arguments 2* and *4* would be disquotation-first viewpoint-arguments.

**Assessments and semantic sub-values.** We can now give a more precise account of the *assessments* and *semantic subvalues*. An assessment is an assessment of an atomic sentence *S* at a viewpoint. It is based on the conclusion of a viewpoint-argument at that

---

<sup>14</sup> This assessment sensitivity is unsavoury, since whether something is true or false should not depend on which (compelling) argument one uses to obtain a truth-value ascription to it. I am not at all concerned with assessment sensitivity that is not of this kind and suspend judgment on theories that claim that assessment sensitivity is a semantic fact for epistemic modals, expressions of personal taste and the like.

<sup>15</sup> In his defense of the assessment sensitivity of liar sentences, Alexandre Billon suggests that, in the case of sentence types, viewpoints correspond to arguments (Billon 2011). So, in my view, he gets it almost right. But why arguments? Well, empirical data, perception, and intuition seem not very promising starting points for liar and truth-teller sentences.

<sup>16</sup> The viewpoint arguments are not part of the natural language fragment  $\mathcal{L}_T$ .

viewpoint. Each such conclusion assigns a semantic value to *S*. These values are relative to the viewpoint. As such they cannot be the semantic values of truth or falsehood *tout court*. Rather, they are the *semantic sub-values* that are constitutive for those values. There are exactly two such values: TRUE (i.e. at a viewpoint) and FALSE (i.e. at a viewpoint). The assessments are thus bivalent.

**How viewpoints are related to viewpoints.** A viewpoint is defined not just by its viewpoint arguments, but also by what other viewpoints it has access to. (I only consider relations in which the context does not change from one viewpoint to the next.) First, naturally, any viewpoint is such that if someone has taken it, then they have access to it (reflexivity). Second, some viewpoints are such that someone who has taken them can, from them, take another viewpoint. And since once they have taken that other viewpoint they would be able to take whatever viewpoints *that* viewpoint can take (if any), such a person would also have access to those viewpoints (transitivity). There are also viewpoints such that someone who has taken them cannot take another viewpoint. Such viewpoints can be described as satisficer viewpoints. A satisficer is *disposed to cease looking* for alternatives once they have found what they were looking for—in our case a per se irrefutable argument with a conclusion that provides a semantic-value ascription to the input sentence. Nonsatisficers are *disposed to continue looking* for an alternative viewpoint. So, with regard to (L), from a non-satisficer viewpoint someone could have access to *arguments 1* and *2*, or to *arguments 3* and *4*, but from a satisficer viewpoint, they would only have access to one of these arguments each time. Both satisficer and nonsatisficer viewpoints are (86) entirely *rational* viewpoints to hold. We don't usually require someone to produce more than one valid and per se irrefutable argument to prove a conclusion, but we also usually don't object when someone produces more than one such argument.

Using *arguments 1* to *4* from above, one can see that, even with their context fixed, ***liar sentences are assessment sensitive***, since their semantic subvalue depends upon the viewpoint from which they are assessed. Truth-teller sentences do not emerge as assessment sensitive, as one can see by using *argument forms 1* to *4* for (T) rather than (L). If one widens the scope of assessments to other atomic sentences, including semantic-value-ascribing non-self-referential sentences, other self-referential sentences, and those that are neither, none of these result in being assessment sensitive either. Liar sentences are special in this regard.

Based on the argument-based assessments at viewpoints, we can now define the semantic values of atomic sentences and their interrelations. If, with its context fixed, a sentence  $S$  has the same subvalue regardless of viewpoint, I say  $S$  has the value **true-regardless** (or **false-regardless**), because its value is regardless of viewpoint. Otherwise, I say  $S$  has the value **true-dependent**, since its value depends on the viewpoint taken. For obvious reasons, we have

- (5) Whenever a sentence ‘ $\ulcorner S \urcorner$ ’ is true-dependent, its negation is also true-dependent.<sup>17</sup>

We could also say that whatever is true-dependent is false-dependent as well. For different obvious reasons, we also have

- (6) Whenever a sentence ‘ $\ulcorner S \urcorner$ ’ is true-regardless, then  $\ulcorner \neg S \urcorner$ .<sup>18</sup>

Since truth and falsehood (*tout court*) should not depend on what per se irrefutable argument we use to infer a sentence’s truth value, I identify truth-regardless with *truth (tout court)* and falsehood-regardless with *falsehood (tout court)* in a natural language like English. In line with section 4.1, I use the sentential operator ‘it is true that’ to express truth-regardless and use ‘it is true that (it is) not (the case that)’ to express falsehood-regardless. (There are clear similarities to the notions of stable, or nearly stable, truth and falsehood and unstable values, in revision theories of truth such as Herzberger 1982a; and Belnap and Gupta 1993.)

---

<sup>17</sup>  $S$  is true-dependent whenever it is neither true-regardless nor false-regardless. Its negation  $\neg S$  is true-dependent whenever  $it$  is neither true-regardless nor false-regardless. Whenever there is a viewpoint from which  $S$  is false-regardless,  $S$  is not true-regardless and  $\neg S$  is not false-regardless. Whenever there is a viewpoint from which  $S$  is true-regardless,  $S$  is not false-regardless and  $\neg S$  is not true-regardless. So, whenever there are both a viewpoint from which  $S$  is true-regardless and a viewpoint from which  $S$  is false-regardless, there are also both a viewpoint from which  $\neg S$  is true-regardless and a viewpoint from which  $\neg S$  is false-regardless.

<sup>18</sup> To show (6), suppose: ‘ $\ulcorner S \urcorner$ ’ is true-regardless and it is not the case that  $\ulcorner S \urcorner$ . Then things are not as ‘ $\ulcorner S \urcorner$ ’ says. Then ‘‘ $\ulcorner S \urcorner$  is false’ is the conclusion at a viewpoint (see above). Then ‘ $\ulcorner S \urcorner$ ’ has the subvalue FALSE at that viewpoint. Then ‘ $\ulcorner S \urcorner$ ’ is not true-regardless.

So, (L) is assessment sensitive and true-depending, regardless of context. (T) is assessment insensitive while being constantly true-regardless in one context and false-regardless in the other. Gestalt shifts occur in disquotation-first arguments in both cases. (They are required for the recognition that after the use of disquotation, we have a truth-value ascription to the sentence – either when moving (87) from line 1 to line 2 or when contemplating line 2.) In truth-teller sentences the Gestalt shifts are benign. In liar sentences they are malignant. The paradoxicality of liar sentences results from the combination of a pragmatic element, i.e., their bistability, with the self-ascription of a semantic value that is incompatible with truth, a combination that leads to the (bivalence-undermining) assessment sensitivity of liar sentences.

#### **4.6. The Undecidable Semantic Status of Liar Sentences and Liar Agnosticism**

One would be misjudging the nature of the assessment sensitivity laid out in the previous section, if at this point one were to infer that it is my suggestion that liar sentences are neither true nor false, or are both true and false, or have no semantic value. Any such further step of introducing an *absolute* semantic status for liar sentences would be utterly misguided. Rather, my point is precisely that the *relative*, viewpoint-dependent, semantic subvalues of liar sentences are as far as one can get with regard to the semantic status of liar sentences.

It is by recognizing and acknowledging this fact, that one finds the appropriate epistemic stance towards liar sentences. What is specific about the viewpoint-arguments is that *each one of them is fully legitimate and per se irrefutable*. The conclusion of each viewpoint-argument is as justified as any further conclusion to the effect that a sentence's semantic value depends on one's viewpoint. For illustration: the conclusion that (L) is true (by *argument 1* or *4*) is as legitimate as the conclusion that (L) is false (by *argument 2* or *3*) and is as legitimate as any further conclusion that at one viewpoint (L) is true, at another false, i.e., that (L) is true-depending. Each of these three conclusions results from flawless reasoning. (In the first two cases, if someone looks no further, we have a satisficer viewpoint, in the last case we have a nonsatisficer viewpoint.) We do not expect someone who produces the irrefutable argument that concludes that (L) is true to go looking for alternative arguments. Nobody can

be faulted for adhering to their conclusion.<sup>19</sup> In other words, *it is characteristic of the assessment sensitivity that institutes the semantic status of liar sentences that it itself is assessment sensitive*. It is because of this, that the appropriate epistemic attitude toward *any* semantic status of liar sentences is suspension of judgement—with regard to their truth or falsehood as much as with regard to their assessment sensitivity. The appropriate reaction is *liar agnosticism*.

Let me explain this a little more formally. For purposes of illustration, I fix (L)’s context as  $c_1$  ‘things are as (L) says’. With  $c_1$ , for rational individuals, there could in principle be the following four viewpoints: the two satisficer viewpoints that, respectively, house *argument 1* or *argument 2* from the previous section, and that have no access to other viewpoints; and two nonsatisficer viewpoints. One of these would house *argument 1* and have access to a viewpoint housing (88) *argument 2*. The other would house *argument 2* and have access to a viewpoint housing *argument 1*.<sup>20</sup> I call these four viewpoints viewpoints 1, 2, 3, and 4, in this order. Given  $c_1$ , someone at one of the two nonsatisficer viewpoints can make the following further inference from their two viewpoint-arguments:

It depends on one’s viewpoint whether (L) is true or false.

Or, what is the same,

(L) is true-dependent.

So, with the context of evaluation ‘things are as (L) says’ selected, there are three conclusions provided by the four possible viewpoints. Each appears to have been reached by impeccable reasoning. Analogous results are obtained with the context  $c_2$ , ‘things are not as (L) says’. For contrast, with (T) only different contexts of evaluation provide different truth values.

---

<sup>19</sup> Thus, among rational viewpoints to hold, there always exist satisficer viewpoints with the viewpoint-arguments 1 and 2.

<sup>20</sup> This is only a partial description of rational viewpoints. There are two options for nonsatisficer viewpoints. The viewpoint they access may in turn be a satisficer viewpoint or a nonsatisficer viewpoint. So, in the first nonsatisficer case, the viewpoint housing *argument 2* could be a satisficer or a nonsatisficer viewpoint, and in the second case, the viewpoint housing *argument 1* could be. Since among the (rational) viewpoints there always is a satisficer viewpoint, each nonsatisficer viewpoint either directly or indirectly accesses a satisficer viewpoint.

The outcome concerning (L) points to a crucial feature of truth. The argument housed at the satisficer viewpoint 1 concludes that (L) is true. It results from an irrefutable argument. Nothing suggests the conclusion is come by dishonestly. We do not expect someone who produces this kind of argument to go look for further or alternative arguments. That is, from viewpoint 1, (L) is true-regardless. Mutatis mutandis the same holds for the satisficer viewpoint 2. From viewpoint 2 (L) is false-regardless. So, with the context fixed, both the value true-regardless and the value false-regardless of (L) can be rationally defended on their own. The third possible scenario is that someone takes a nonsatisficer viewpoint, i.e., a viewpoint that takes an alternative viewpoint, if there is one. Viewpoints 3 and 4 are examples. In such a scenario, effectively the overall result is that (L) is true-dependent. Since there is nothing irrational about someone searching for more than one argument that produces a semantic value for (L), this result, too, can be rationally defended. There are then *equally justifiable* viewpoints that result in (L) being true-regardless, false-regardless, and true-dependent. Consequently, it depends on what viewpoint one takes, whether (L) is assessment sensitive. In other words, it is true-dependent whether (L) is true-dependent.<sup>21</sup>

One can show that *this* result is true-dependent as well. All one needs to concede is that the holder of a rational viewpoint can access and assess (by further argument) their previously reached results regarding (L) and that truth-regardless and falsehood-regardless are luminous. Both are points that are generally granted. Then the following principle holds.

(7) If some sentence is true-regardless or false-regardless, then a sentence expressing that this is so is true-regardless. (89)

With (7), for viewpoint 1 we obtain the argument

---

<sup>21</sup> Objection: But if one can take a viewpoint which accesses two arguments that come to conflicting conclusions, then surely each of these arguments alone is no longer defensible. Reply: This is not so. Given that the context is fixed, the arguments are each flawless. The facts that, looked at differently, a different conclusion results and that, looked at in two ways, two conflicting conclusions result do not change the fact that, as it is, the one argument is legitimate and irrefutable and nobody can be faulted for adhering to its conclusion, no matter what. Arguing otherwise is basically simply denying the paradoxicality of the Liar. At the minimum, a flaw in the argument would need to be identified. The fact that, looked at differently, a different conclusion results is not *an indication that the argument is flawed*. Instead, it may be an indication that something is not quite right with (L).



1. (L) is true-regardless. Viewpoint 1
2. ‘(L) is true-regardless’ is true-regardless. (7)

And for viewpoint 2 we obtain the argument

1. (L) is false-regardless. Viewpoint 2
2. ‘(L) is false-regardless’ is true-regardless. (7)

These satisficer viewpoint-arguments, too, are flawless. Moreover, at each viewpoint such arguments using (5.1) can be repeated ad infinitum, each time applying (5.1) to the conclusion of the preceding argument.

So, since there are (i) viewpoints that result in the truth-regardless of the truth-regardless of (L) and (ii) viewpoints that result in the truth-regardless of the falsehood-regardless of (L) being false-regardless, and (iii) viewpoints that result in the truth-dependence of the truth-dependence of (L) (see above), we can infer that it is true-dependence that it is true-dependence that (L). Analogous arguments of this type can be developed to obtain further iterations of ‘truth-dependence’. It is thus a characteristic of truth-regardless that

- (8) If a sentence is viewpoint-dependent regarding its semantic status, then it is viewpoint-dependent whether this is so (i.e. whether the sentence is viewpoint-dependent regarding its semantic status), and viewpoint-dependent whether this in turn is so, and so on.

This means that, if we take seriously the fact that the viewpoint-arguments housed in the various viewpoints are impeccable—and we have no reason not to, since they are per se irrefutable—the kind of agnosticism it is rational to hold about the Liar is the following. Not only do we need to suspend judgement w.r.t. whether liar sentences have truth or falsehood as their semantic value. We also need to suspend judgement w.r.t. whether liar sentences are viewpoint-dependent, and w.r.t. whether it is viewpoint-dependent whether liar sentences are viewpoint-dependent, and so on. Elsewhere, I call such agnosticism *absolute* agnosticism

(Bobzien 2010). I propose that the epistemic stance of absolute agnosticism is the correct one for the Liar.

#### **4.7. The Normal Modal System KT4M with the Truth Operator T for Truth-Regardless**

In the next two sections I provide a formalization of the truth-operator. In this section I provide a formal representation of the notion of truth (p.90) that complements liar agnosticism. In section 4.8, I add a matching model-theoretic semantics. In Section 4.9, and based on the formal representation and model-theoretic semantics, I introduce a procedure that makes it possible to show of non-paradoxical sentences that they are bivalent. The philosophical argument continues in section 4.10.

Any formal representation of truth in natural languages needs to consider whether to represent something's being true by means of a truth predicate or a truth-operator. I use a sentential operator 'it is true that', for reasons set out in section 4.1. In section 4.5, I identified truth with truth-regardless and falsehood with false-regardless as set out in terms of the sub-values TRUE and FALSE in that section. The value-ascribing predicates 'is true-regardless' and 'is false-regardless' are related to the truth-operator as follows (where 'rS' is the name of the sentence abbreviated as S):

(9) A sentence 'rS' is true-regardless iff it is true that S.

(10) A sentence 'rS' is false-regardless iff it is true that [it is] not [the case that] S.

('It is false that S' and 'It is true that [it is] not [the case that] S' are assumed to be equivalent.) In common understanding, a sentence is bivalent if it is either true or false. In terms of truth-regardless, a sentence is defined as bivalent if it is either true-regardless or false-regardless, and that is, if it is not true-dependent.

(11) A sentence 'rS' is bivalent iff either it is true that S or it is true that [it is] not [the case that] S.

I symbolize the sentential operator ‘it is true that’ with an italicized bold capital  $T$ . Syntactically,  $T$  is modeled on the necessity operator  $\Box$  in normal modal logics and can be iterated indefinitely. From classical logic I add  $p, p_1, p_2, \dots$  for atomic sentences and the connectives  $\neg, \wedge, \vee, \rightarrow$  and  $\leftrightarrow$ . I adopt the common method of describing modal systems with schemata of axioms and theorems and henceforth use ‘axiom’ as short for ‘axiom schema’ and ‘theorem’ as short for ‘theorem schema’.  $A, A_1, A_2, \dots$  are metalinguistic expressions for arbitrary well-formed formulas.

The results of sections 4.5 and 4.6 suggest that the operator  $T$  is governed by the principles and rules that govern the modal system KT4 (that is, the rules PC, that all tautologies of propositional calculus are axioms, MP (*modus ponens*) and N (Necessitation), and the axioms K, T, 4) plus *one additional axiom*. The rules MP, N, and the truth axioms  $K_T$  ( $[TA_1 \wedge [TA_1 \rightarrow TA_2]] \rightarrow TA_2$ ),  $T_T$  ( $TA \rightarrow A$ ) and  $4_T$  ( $TA \rightarrow TTA$ ) should be uncontroversial. Moves to drop any of them *in order to* solve the liar paradox are here disregarded, for obvious reasons. The same holds for dropping PC. Axiom  $4_T$  expresses the luminosity of ‘it is true that’ that was introduced informally as (7) in section 4.6. Axiom  $T_T$  expresses (6) from section 4.5. It is the modal correlate to semantic descent.

For convenience, I define an operator that expresses truth-dependent (‘it is not true that  $A$  and it is not false that  $A$ ’) using the symbol  $\mathcal{F}$ .

$$\mathcal{F}A =_{\text{def}} \neg TA \wedge \neg T\neg A.$$

( $\mathcal{F}$  can be visualized as a combination of the letters T and F—a reminder that  $\mathcal{F}A$  results from truth at one viewpoint and falsehood at another.)  $\mathcal{F}$  parallels the downward-pointing triangle  $\nabla$  (‘it is contingent that’) from contingency logic. The informal principle (5) from section 4.5 can then be formally expressed as the truth-correlate to the Mirror Theorem, i.e., as  $\mathcal{F}A \leftrightarrow \mathcal{F}\neg A$ . The operator  $\mathcal{F}$  also provides a succinct way of modally expressing bivalence.  $\neg \mathcal{F}A$  expresses that either it is true that  $A$  or it is false that  $A$ . By (11), this provides the modal correlate to the bivalence of  $A$ .

*The additional axiom* is designed to capture the informal principle (8) from section 4.6 that motivates suspension of judgement on liar sentences. This principle generalizes to ‘when it is true-dependent whether  $A$ , then it is true-dependent whether it is true-dependent whether  $A$ ’, formally

$$M_{\mathcal{F}} \quad \vdash \quad \mathcal{F}A \rightarrow \mathcal{F}\mathcal{F}A$$

or, in terms of the  $T$ -operator,

$$[\neg TA \wedge \neg T\neg A] \rightarrow [\neg T[\neg TA \wedge \neg T\neg A] \wedge \neg T\neg[\neg TA \wedge \neg T\neg A]].$$

System  $KT4_T$  supplemented by axiom  $M_T$  provides the formal equivalent to the representation of truth in natural language<sup>22</sup> as developed in sections 4.1 to 4.6. In system  $KT$ , and hence in  $KT4$ , the triangle equivalent to  $M_T$  (i.e.  $\nabla A \rightarrow \nabla \nabla A$ ) is equivalent to the McKinsey axiom

$$M \quad \vdash \quad \Box \Diamond A \rightarrow \Diamond \Box A.$$

This can be easily demonstrated.<sup>23</sup> Accordingly, the proposed formal representation of truth in natural language is structurally equivalent to the normal modal system  $KT4M$ .<sup>24</sup> It has a truth operator that is governed by a normal modal logic and has a Kripke semantics, or possible-worlds semantics. *It is thus coherent.* Henceforth, ‘ $KT4M$ ’ is understood as denoting system  $KT4M$  with the truth-operator  $T$  instead of the  $\Box$ -operator. There are clear parallels in the use of  $KT4$  for a truth-operator (i) in predicative axiomatic theories of truth, where Kripke-Feferman (Feferman 1984) is very close, with its axioms  $KF1-5$  and  $KF8-10$  (following Horsten 2011), and (ii) in revisionist theories of truth, where Herzberger’s schemata i to iv are similarly close (Herzberger 1982a, 495–96). Still, for obvious reasons, (92) neither Kripke-Feferman nor Herzberger contains an unrestricted necessitation rule (more on which in section 4.10, where I discuss the ‘liar property’) and neither contains axiom  $M$  or the relation of finality. With the falsehood of  $A$  formalized as  $T\neg A$ , by means of axioms  $T$  and 4,  $KT4M$  also provides roughly functional analogues (in modal terms) to the Kripkean fixed points:  $TA \leftrightarrow T^n A$  for any  $n$ , for truth and  $T\neg A \leftrightarrow T^n \neg A$  for any  $n$ , for falsehood.

#### 4.8. A Modal Semantics for $KT4M$ and Its Representation of Truth (i.e. Truth-Regardless)

In this section I provide a model-theoretic Kripke semantics for  $KT4M$  as the *logic of truth in natural language*, with an interpretation that reflects the context sensitivity of (L) and (T) and the assessment sensitivity of (L) as discussed informally in earlier sections.

<sup>22</sup> Sections 4.12 to 4.14 specify further how truth in natural language is represented.

<sup>23</sup> E.g. Bobzien (2015), pp. 85–85.

<sup>24</sup>  $KT4M$  is also known as  $S4M$  and as  $S4.1$ . For some historical background see e.g. Hughes and Cresswell (1996). For its decidability see Segerberg (1968).

Here are the customary basics of the modal semantics KT4M.  $p, p_1, p_2, \dots$  are countably infinitely many *atomic sentence formulas*. A *frame*  $\mathcal{F}$  is an ordered pair  $\langle W, R \rangle$ , where  $W$  is a nonempty set of objects ('points') and  $R$  is a binary accessibility relation defined over the members of  $W$ , so that it is determinate for any points  $w, w'$  in  $W$  whether  $wRw'$ . A *model*  $\mathcal{M}$  is an ordered triple  $\langle W, R, V \rangle$  where  $W$  is a nonempty set of objects ('points'),  $R$  is a binary *accessibility relation* defined over the members of  $W$ , and  $V$  is a *value assignment* for the atomic sentence formulas of  $\mathcal{M}$ , satisfying the standard value assignment rules (R1) to (R3):

(R1) For any  $p$  and any  $w \in W$ , either  $V(p, w) = 1$  or  $V(p, w) = 0$ .

(R2)  $[V\neg]$  For any wff  $A$ , and any  $w \in W$ ,  $V(\neg A, w) = 1$  if  $V(A, w) = 0$ ; otherwise,  $V(\neg A, w) = 0$ .

(R3)  $[V\vee]$  For any wff  $A$  and  $B$ , and for any  $w \in W$ ,  $V((A \vee B), w) = 1$  if either  $V(A, w) = 1$  or  $V(B, w) = 1$ ; otherwise,  $V(A, w) = 0$ .

Instead of the rule for the necessity operator  $\Box$  there is an analogous one for the truth-operator  $T$ :

(R4)  $[VT]$  For any wff  $A$  and for any  $w \in W$ ,  $V(TA, w) = 1$  if for every  $w' \in W$  such that  $wRw'$ ,  $V(A, w') = 1$ ; otherwise  $V(TA, w) = 0$ .<sup>25</sup>

Conditions for the sentential operators  $\wedge, \rightarrow, \leftrightarrow$ , and  $F$  can be derived from the above in the usual way. In terms of frames, the system KT4M can be characterized by the class of frames that are transitive and reflexive and in which every world can access at least one world that can access only itself. This last condition is known as *finality*.<sup>26</sup>

In line with sections 4.1 to 4.7, one can produce the following model-theoretic representation of *truth-regardless* based on the class of KT4M frames. Its purpose (93) is (i) to provide a formal correlate to the informally introduced—liar-accommodating—notion of *truth-regardless*, (ii) to show how the structural elements of *truth-regardless* can be encoded in such a model. and (iii) to prove the coherence of the notion of *truth-regardless*. The basic parameters of the representation are these:

---

<sup>25</sup> For a rebuttal of the—misguided—objection that we do not need a possible-world semantics for truth, since there is a truth evaluation 'built into' possible worlds see Leitgeb (2003, p. 129).

<sup>26</sup> For soundness and completeness proofs for KT4M see, e.g., Hughes and Cresswell (1996). Finality has been shunned in theories of truth, though it is discussed in a paper whose authors, like myself, believe that truth and necessity should be approached in similar ways—if they do it the other way about than I (Halbach, Leitgeb, and Welch 2005).

- *Atomic sentences* of a natural language like English  $\mathcal{L}$  (including atomic liar and truth-teller sentences).
- *Standard conditions of value assignments for complex sentences formed with the operators*  $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ .
- *Assessment viewpoints* for the sentences. For each atomic sentence, such a viewpoint houses a per se irrefutable argument with a conclusion that is a semantic-value ascription to that sentence. Dispositionally, each viewpoint is either a satisficer viewpoint or a nonsatisficer viewpoint (section 4.5).
- The *semantic subvalues* TRUE and FALSE.
- *Argument-based assessments* that assign a semantic subvalue to each sentence at each viewpoint.
- The *subvalues for atomic sentences* are obtained as set out in section 4.5. Subvalues for nonatomic sentences are obtained in accordance with Rules (R2) to (R4), where (R2) governs natural-language negation, (R3) natural-language binary inclusive disjunction, and (R4) natural-language sentential-operator use of ‘true’ and ‘false’ each understood as viewpoint-independent.
- *A reflexive, transitive, and final binary accessibility relation of being disposed to take a viewpoint from a viewpoint*, based on section 4.5. (Being disposed to take a viewpoint is taken to entail being able to take that viewpoint.)
- Truth-regardless, truth-depending, and bivalence as *the resultant truth-modalities* of the sentences.
- *Circumstances*. These ensure that the context of the liar and truth-teller sentences is fixed.

The correlation between the model-theoretic framework and the parameters of its representation is the following. The **atomic formulas**  $p, p_1, p_2, \dots$  encode logically independent atomic sentences of  $\mathcal{L}$ . The **values** **1** and **0** encode the semantic subvalues TRUE and FALSE.<sup>27</sup> Each **point**  $w$  encodes a possible assessment viewpoint for its set of atomic sentences  $p_1$  to  $p_n$ . The **set of points**  $\mathbf{W}$  corresponds to a nonempty set of possible viewpoints at which all sentences of a frame are assessed.  $\mathbf{V}(p, w) = \mathbf{1}$  encodes that, assessed at the

---

<sup>27</sup> For the relation between TRUTH and truth *simpliciter* see section 4.5. My subvalues TRUE and FALSE would correspond to Leitgeb’s *internal truth*, and the truth expressed by the truth-operator to his *external truth* (Leitgeb 2003, p. 129), though otherwise what I offer is very dissimilar to his approach.

viewpoint  $w$ ,  $p$  has the subvalue TRUE.  $\mathbf{V}(p, w) = \mathbf{0}$  encodes that, assessed at  $w$ ,  $p$  has the subvalue FALSE. Thus ‘truth-at-a-point- $w$  of  $A$ ’ is a semantic value that is relative to the assessment of  $A$  at  $w$ , that is, TRUTH. The reflexive, transitive and final **accessibility relation**  $\mathbf{R}$  encodes the disposition to take a viewpoint from a viewpoint. So  $wRw'$  is interpreted as the relation of one being disposed to take the viewpoint  $w'$  from one’s viewpoint  $w$ . Reflexivity, transitivity, and finality of (94) the relation are fleshed out thus. **Reflexivity**: Each viewpoint can be taken from itself. Someone who has a viewpoint is understood to be disposed to take it. **Transitivity**: Being disposed to take a viewpoint is the same as being disposed to get to have that viewpoint. Accordingly, one can take every viewpoint *that* viewpoint can take. **Finality**: Every viewpoint is disposed to take a viewpoint that is only disposed to take itself. That is, by ‘collecting’ viewpoints via the relation  $\mathbf{R}$ , from every viewpoint one may start out from, one gets to a viewpoint that opens up no further viewpoints. Such a viewpoint is a *final viewpoint*. Someone at that viewpoint lacks the disposition to take any other viewpoint. Given reflexivity, final viewpoints are satisficer viewpoints and vice versa. Given reflexivity, transitivity, and finality, every nonsatisficer viewpoint is disposed to take a satisficer viewpoint. A **frame**  $\mathcal{F}$  encodes pathways of collecting viewpoints with regard to its atomic sentences  $p_1$  to  $p_n$ . A **model**  $\mathcal{M}$  encodes a possible subvalue assignment over  $\mathcal{F}$  to the atomic sentences  $p_1$  to  $p_n$  of the viewpoints of  $\mathcal{F}$ . The **valuated set of points**  $\mathbf{W}$  in a model  $\mathcal{M}$  encodes a nonempty set of viewpoints that each comprises an assessment, i.e., subvaluations for each of its atomic sentences.

Finally, circumstances. In order for liar and truth-teller sentences to be semantically evaluable, their context needs to be fixed (section 4.4). This requirement is satisfied in the representation of truth in KT4M via circumstances. Note that the conditions for fixing the context of the subject expressions of liar and truth-teller sentences from section 4.5 are identical with the following conditions: (i) Things are as (L) ((T), etc.) says. (ii) Things aren’t as (L) ((T), etc.) says. I call such conditions circumstance-conjuncts of a sentence. As it happens, every atomic sentence of  $\mathcal{L}_T$ —whether context sensitive in the manner of (L) and (T) or not—has two such circumstance-conjuncts. The semantic evaluability of liar and truth-teller sentences is then ensured in  $\text{KT4M}_T$ , if in every model over every frame each atomic sentence is mapped onto precisely one of its two circumstance-conjuncts, for example  $p_1$  on (i),  $p_2$  on (ii),  $p_3$  on (ii), etc. I call the conjunction of those circumstance-conjuncts of a model  $\mathcal{M}$  the

*circumstance of  $\mathcal{M}$* . In all nontrivial cases there will be a plurality of models with the same circumstance.<sup>28</sup>

*Validity and consistency* of sentences in  $\text{KT4M}_T$  are defined in the standard way. A sentence  $A$  is *valid* in  $\text{KT4M}_T$  precisely if it is valid on all frames of  $\text{KT4M}_T$ . A sentence  $A$  is *consistent* in  $\text{KT4M}_T$  precisely if there is a viewpoint  $w$  in a model  $\mathcal{M}$  of a frame  $\mathcal{F}$  of  $\text{KT4M}_T$  such that  $A$  is TRUE at  $w$ .

This concludes the basic representation of truth-regardless and truth-depending by means of  $\text{KT4M}$ . Since  $\text{KT4M}$  is a normal modal logic, the notion of truth as truth-regardless developed in sections 4.1 to 4.6 and represented as semantic necessity, symbolized by  $T$ , is coherent within a classical, bivalent logic.

#### **4.9. Adding to $\text{KT4M}$ the Distinction between Assessment-Sensitivity and Assessment-Insensitivity**

For  $\text{KT4M}$  to do more than define the structure of truth, we need to single out individual atomic sentences as assessment-insensitive.

In the semantics, this requires a *partial interpretation* of  $\text{KT4M}$  coupled with a *constraint on the class of models*. This constrained  $\text{KT4M}$  or C- $\text{KT4M}$  will also further elucidate how my usage of Kripke semantics differs from uses as possible-world semantics that take points to be worlds. The relevant partial interpretation incorporates into  $\text{KT4M}$  the distinction between assessment-sensitive and assessment-insensitive sentences from Section 4. Each atomic sentence  $p_1, p_2, \dots$  in each model is interpreted either as being assessment-sensitive or as being assessment-insensitive. The class of models is then restricted to a subclass of models in which the same atomic sentences are marked out as assessment-insensitive and in which not all atomic sentences are marked out as assessment-insensitive. I shall refer to this partially interpreted and constrained system as C- $\text{KT4M}$ . With the philosophical results from Sections 2 to 5, a sentence  $p$  is semantically non-paradoxical precisely if it is assessment-insensitive, and a sentence  $p$  is semantically paradoxical precisely if it is assessment-sensitive. I take the two cases in turn.

---

<sup>28</sup> Circumstances can perhaps be elucidated by comparison with possible-world semantics that have designated actual worlds. Circumstances are somewhat similar to whatever it is regarding a model that makes the actual world the actual world in that model (perhaps truth conditions of the actual world of  $\mathcal{M}$  in such a semantics). Yet, in the representation of truth by  $\text{KT4M}$ , there are no designated actual worlds.



**(i) Atomic assessment-insensitive sentences.** With its context fixed, any assessment-insensitive sentence  $p$  has by definition the same conclusion in its arguments at all viewpoints. This conclusion is either that  $p$  is false or that  $p$  is true. In terms of the semantics of KT4M, each interpreted atomic sentence  $p$  has as its circumstance-conjunct either that things are as  $p$  says or that things aren't as  $p$  says. This restricts the viewpoint-arguments for assessment-insensitive  $p$  at each viewpoint to those that have the corresponding sentence ('things are as  $p$  says', 'things aren't as  $p$  says') as a premise. (Or in any event I restrict the arguments thus, since otherwise there would be an inconsistency between the circumstance-conjunct and the first lines of the arguments.) When  $p$ 's circumstance-conjunct in a model  $\mathcal{M}$  is that things are as  $p$  says, then no per se irrefutable argument is possible with the premise 'things aren't as  $p$  says'. When  $p$ 's circumstance-conjunct in the model  $\mathcal{M}$  is that things aren't as  $p$  says, then no per se irrefutable argument is possible with the premise 'things are as  $p$  says'. All viewpoint-arguments for  $p$  in  $\mathcal{M}$  have one or the other as premise. As a result, in any model  $\mathcal{M}$  the semantic subvalue ascribed to an assessment-insensitive  $p$  will be constant across viewpoints: constantly TRUE or constantly FALSE.

One can then show as follows that in C-KT4M all semantically nonparadoxical sentences are bivalent, or, what is the same, that  $\mathbf{T}p \vee \mathbf{T}\neg p$  holds of every nonparadoxical sentence. In the semantics of C-KT4M, in any model  $\mathcal{M}$  over any frame  $\mathfrak{F}$ , (96) assessment-insensitive sentences have constant subvalues across all viewpoints of  $\mathcal{M}$ . The value of  $p$  across all viewpoints of  $\mathcal{M}$  is either true or false. Thus, if  $p$  is nonparadoxical,  $\mathbf{T}p \vee \mathbf{T}\neg p$  is valid<sub>on- $\mathfrak{F}$</sub>  for any  $\mathfrak{F}$  of the class of frames of C-KT4M. Hence, by the definition of validity, for any nonparadoxical atomic  $p$ ,  $\mathbf{T}p \vee \mathbf{T}\neg p$  is valid *tout court* in C-KT4M. Thus, in modal terms, in C-KT4M *all semantically nonparadoxical atomic sentences are bivalent*.

**(ii) Atomic assessment-sensitive sentences.** assessment-sensitive sentences are true-depending. With its context fixed, any assessment-sensitive atomic sentence  $p$  has, by definition, a viewpoint at which its argument concludes that  $p$  is true and another viewpoint at which its argument concludes that  $p$  is false. In terms of the semantics of C-KT4M, again, in any model  $\mathcal{M}$ , each interpreted atomic sentence  $p$  has as its circumstance-conjunct either that things are as  $p$  says or that things aren't as  $p$  says. But this time there will be at least one model  $\mathcal{M}$  with one viewpoint at which  $p$  is true and another viewpoint at which  $p$  is false. By (6), the same holds for the negations of assessment-sensitive atomic sentences. If we apply C-KT4M to such assessment-sensitive sentences, we immediately obtain the following results.

*No atomic assessment-sensitive sentence  $p$  or its negation is valid in C-KT4M and of no assessment sensitive sentence  $p$  is any  $\mathbf{T}^m\mathbf{F}^n p$  or any  $\mathbf{T}^m\mathbf{F}^n \neg p$  (with  $m \geq 0$  and  $n \geq 1$ ) valid in C-KT4M.*

Now, if a sentence is not valid in a logical system, then the system itself does not provide sufficient reasons for holding that sentence.<sup>29</sup> Thus C-KT4M does not provide sufficient reason for holding any sentence that is assessment sensitive; nor for the holding of any sentence, including assessment sensitive ones, that it is true-dependent; nor for the holding of any sentence, including assessment-sensitive ones, that it is truly true-dependent, truly truly true-dependent, etc. Since liar sentences and their kin are the only candidates for assessment sensitivity, C-KT4M accurately delivers the result that absolute agnosticism regarding the semantic status of liar sentences, as presented in section 4.7, is the appropriate epistemic stance to adopt about the liar paradox. *For liar sentences, bivalence can neither be validated nor invalidated.*

#### **4.10. The Liar Property, Gestalt Shifts, KT4M, and the Solution**

From the definitions of ‘true-dependent’ and ‘consistency’ it follows that if a sentence is true-dependent ( $\mathbf{T}$ ) in KT4M, it is consistent in KT4M. So, insofar as liar sentences are true-dependent, in KT4M they are consistent. ‘But what about the liar property?’ some are bound to ask at this point. What (97) about the liar property? By this I mean the property commonly expressed, with predicative semantic-value ascription, in a biconditional, sometimes relative to a language  $\mathcal{L}$ , as a semantic relation as follows:

$$(12) \quad (\mathbf{L}) \leftrightarrow \text{‘}(\mathbf{L})\text{’ is false.}^{35}$$

(For the version with ‘untrue’ for ‘false’ see section 4.11.) The modal syntax of KT4M would allow us to express the liar sentence (L) both as L and as  $\mathbf{T}\neg\mathbf{L}$ . (I do not consider the round brackets in (L) as part of the name of the liar sentence and omit them in modal formalizations, to prevent readers from inadvertently parsing the formulas as predicate formulas.) Also,

---

<sup>29</sup> Plainly, I am not talking probabilistic logic here.

<sup>30</sup> I disregard the relation between the liar property and diagonalization. How the Diagonal Lemma and Incompleteness Theorem(s) are related to KT4M is the topic of a separate paper. For present purposes, it suffices to take ‘(L)’ to be the name of (L). See also footnotes 2 and 37.

trivially, biconditionals can be expressed in KT4M. So in principle the modal correlate to the liar biconditional (12) could be expressed relative to  $\mathcal{L}_T$  (with  $\mathcal{L}_T$  being the semantically closed propositional fragment of a natural language  $\mathcal{L}$  like English with atomic liar and truth-teller sentences) as

$$(LP) \quad L \leftrightarrow T\neg L$$

So much for syntax. What about the property itself that is usually meant to be expressed by the biconditional (12)? The short answer is that the liar property is already accounted for in KT4M. More specifically, it is accounted for by the combination of theorem  $M_T$  with the interpretation of the operator  $T$  as true-depending and a possible-worlds semantics that incorporates both context sensitivity and assessment sensitivity of semantic-value-self-ascribing sentences. Thus there is *no need* for adding to the logic of truth KT4M a biconditional like (LP) to account for the liar property sentence. In fact, it would be *misguided*, if one were to add such a biconditional. I explain why this is so.

First, I do not deny that there is a liar property of a kind such that it might be tempting to express it in ordinary language versions of (12). Rather, I maintain that this property is a *pragmatic* feature of the liar sentence. Now, pragmatic features of sentences are not usually features a logic is meant to *represent*. Accordingly, KT4M does not *represent* it. There is neither a logical nor a semantic relation that sanctions the moves from  $L$  to  $T\neg L$  and from  $T\neg L$  to  $L$ .<sup>37</sup> What is the pragmatic feature that is the liar property? It is the bistability of ( $L$ ). It is the fact that an individual who entertains ( $L$ ) can experience a shift from understanding it in one way to understanding it in another way (where ‘understanding’ is used to indicate a pragmatic feature as introduced in section 4.3). This shift is a Gestalt shift from the designational understanding to the ascriptional understanding, or vice versa (section 2.3).

How is this feature related to (12) or to (LP)? First, it might seem natural to express ( $L$ ), when understood ascriptionally by ‘ $T\neg L$ ’, and when understood designationally as ‘ $L$ ’. ‘ $T\neg L$ ’ and the atomic ‘ $L$ ’ would then indicate different ways of (98) *understanding* the same

---

<sup>31</sup> KT4M does cover certain logical relations between  $L$  and  $T$ , namely those expressible in propositional modal logic. These include  $TL \rightarrow L$ ,  $T\neg L \rightarrow \neg L$ , and  $L \rightarrow \neg T\neg L$ , via axiom T and, if  $L$  has been proved, the move from  $L$  to  $TL$  via N. Because of  $FL$  (which itself can neither be proved nor disproved), this move can never happen, though.

sentence, or would express the same sentence from different viewpoints. The sentence expressed by either formula would be the liar sentence. Just as the two views of the spinning dancer could be taken to represent the same object, i.e., the same dancer, but from different viewpoints, so ‘ $T\neg L$ ’ and ‘ $L$ ’ could be used to express the same sentence, but to express it differently. Assuming a fixed context (circumstance), both the meaning and the extension of the two expressions ‘ $T\neg L$ ’ and ‘ $L$ ’ would then be the same.

However, this is not what I suggest. It is true that just as one could say that the left-spinning dancer *is* the right-spinning dancer, one could say that  $T\neg L$  *is*  $L$ . But just as the expression ‘left-spinning dancer’ is not the same expression as ‘right-spinning dancer’, so ‘ $T\neg L$ ’ is not the same expression as ‘ $L$ ’. They are different sentences. So, *in the logic KT4M*, they denote different objects. Accordingly, the liar property is taken to allow a prima facie pragmatically licit transition from one sentence to the other.

Just as there are things we can infer from the left-spinning dancer and from the right-spinning dancer that are incompatible (you cannot spin left and right at the same time), so there are things we can infer from the-Liar-qua-understood-one-way that are incompatible with the- Liar-qua-understood-the-other-way, for example we can infer  $T\neg L$  from  $T\neg L$  but not from  $L$ , and can infer  $L$  from  $L$  but not from  $T\neg L$ . Thus pragmatic factors are the reason why incompatible things can be derived from ‘ $T\neg L$ ’ and from ‘ $L$ ’.

Now, of course, you are welcome (i) to add pragmatic information to a logical system, and then (ii) to supplement the rules of that system with whatever pragmatic restrictions are consequently required to keep it consistent. What matters, if you do so, is that you are aware of what you are actually doing. For illustration, I return to the ambiguous sentence (TEL) ‘I saw the woman with the telescope’ from section 4.3. Imagine that, in a system of first-order logic, first you add this ambiguous sentence to your stack of sentences and then you supplement your logic with a restrictive rule, to prevent any inconsistencies from being derivable as a result of your adding this sentence.

Let us add more detail. Take (TEL) and first-order logic (FOL). Start with adding the sentences (TEL-I) for the reading (i) of the sentence (the speaker had the telescope) (TEL-her) for the reading (ii) (the woman had the telescope). Introduce the occurrence of a Gestalt shift from one reading to the other during an inference as (TEL-I) $\rightarrow$ (TEL-her) and (TEL-her) $\rightarrow$ (TEL-I) respectively. Remember that the Gestalt shifts are not in the sentences but in the readers. So, it is at this point that you introduce a pragmatic element into your logic. Next

add a biconditional combining the two possible shifts between readings of the sentence and call this biconditional the Telescope property (TELP)

$$(TELP) \quad FOL \oplus (TEL) \quad \vdash \quad (TEL-I) \leftrightarrow (TEL-her)$$

This (99) will allow you to derive inconsistencies in your language. Next, add a restrictive rule to the logic to prevent any inconsistencies. For instance, you could add the rule (RTel) that whenever a Gestalt shift occurs, nothing derived before the Gestalt shift using reading (i) can be employed to derive anything once you have shifted to reading (ii), and vice versa. Use the following telescope argument as example. (Context: you and the woman are 11 meters apart.)

1.	(TEL-I)	assumption 1
2.	I use the telescope	1
3.	(TEL-I) → (TEL-her)	Gestalt shift (discharge assumption 1)
4.	(TEL-her)	1, 3, <i>modus ponens</i>
5.	She uses the telescope	4
6.	Two people more than 10 meters apart cannot use the same telescope at the same time	Telescope Theorem
7.	¬[I use the telescope]	4, 6
8.	I use the telescope ∧ ¬[I use the telescope]	2, 7, ∧-introduction

Given rule (RTel), you cannot use 2 in step 7 to 8, since it was derived before the Gestalt shift. You could also have a fancier rule, say rule (RTel'), that allows you, after a second Gestalt shift, to access what you had derived before the first Gestalt shift, and after a third Gestalt shift, to access what you had derived after the first and before the second Gestalt shift, and so forth. The general method used should be clear. However, by most logicians, what you are doing would be regarded as pretty idiotic. Mostly, logicians prefer the method of disambiguating a sentence *before* they add it to their logic, and then adding both resulting sentences, and hey pronto.

Back to the liar property (LP). What I suggest is that adding the liar property *as* (LP) to KT4M is methodologically similar—and similarly misguided—to adding the ambiguous sentence (TEL) to first-order logic. The main difference from the telescope example is that where (TEL) is ambiguous and allows two readings, (L) is bistable for reasons of salience and allows two understandings. I go through the same steps. The biconditional (LP) would express a pragmatic feature. Each of its conditionals (LP $\rightarrow$ ) and (LP $\leftarrow$ ) is used to indicate the locus of a Gestalt shift, one for each direction.

If you add this pragmatic relation to your logic, you will be able to derive inconsistencies. Let me illustrate this also. As example, I use the following modal representation of a simple Liar argument. Applications of the conditionals (LP $\rightarrow$ ) and (LP $\leftarrow$ ) of the liar property (line 8) are interpreted as representations of Gestalt shifts. (100)

1.	$L \rightarrow T \neg L$	(LP $\rightarrow$ ) (stated)
2.	$T \neg L \rightarrow \neg L$	axiom T
3.	$\neg L \rightarrow \neg TL$	axiom T, contraposition
4.	$L \rightarrow \neg TL$	1, 2, 3, PC
5.	$TL \rightarrow L$	axiom T
6.	$TL \rightarrow \neg L$	4, contraposition, double negation (PC)
7.	$\neg TL$	5, 6, PC
8.	$L$	7, (LP $\leftarrow$ ) (used)
9.	$TL$	8, Necessitation
10.	$TL \wedge \neg TL$	7, 9, $\wedge$ -introduction

Next you add a suitable rule of restriction that reins in the pragmatic element you added, e.g., rule (RLiar) that whenever a Gestalt shift occurs, nothing derived before the Gestalt shift using the designational understanding can be employed to derive anything once you have shifted to the ascriptional understanding, and vice versa. Alternatively, you can add a fancier rule (RLiar') modelled on (RTel'). Both (RLiar) and (RLiar') are rules that require you to revise your set of assumptions and/or derived formulas each time a Gestalt shift occurs, restricting you to those that you obtained with the present Gestalt (or at your present

viewpoint). They are rules of the kind revision theories of truth might employ. With such a revision-theoretical restriction in place, line 4 can no longer be used to derive 7.<sup>32</sup>

Now for the punch line. With the logic KT4M, the introduction of the pragmatically interpreted liar property (LP) followed by pragmatics-based restrictions on the logic is as misguided—if not more so—as the introduction of the telescope-property followed by the rule (RTel) into FOL. To extend the analogy, it is like introducing the telescope-property (TP) together with rule (RTel) into a logic *that already contains* the disambiguated sentences.

On the assumption that the two conditionals of the liar property biconditional chronicle the occurrence of Gestalt shifts, and thus express a pragmatic relation that involves a speaker/listener, an *appropriate* logic for truth would be one in which (a) the inconsistency-introducing pragmatic elements have been eliminated and (b) any relevant informational content of the liar sentence is retained—just as in FOL (a) the ambiguities of sentences are eliminated by disambiguation and (b) the disambiguated sentences have been added.

Now, if we feed the effects of bistability on the liar sentence as viewpoint sensitivity into the possible-worlds semantics, KT4M does exactly that. In particular, it is axiom M that, added to KT4, preempts the need for revision-theoretical rules (or for an N-restriction rule), in the way in which disambiguation preempts the need for a rule like (RTel).

As regards (a), we have seen above that there is no inconsistency in KT4M, and in particular truth is not an inconsistent notion. As regards (b): what is the (101) relevant informational content of the liar sentence that needs to be retained? It is that which belongs to it irrespective of which understanding (designational or ascriptional) someone is entertaining. This would presumably include the fact that it is bistable. If we consider the two expressions L

---

<sup>32</sup> As an alternative to revision-theoretical rules, if you prefer to incorporate the pragmatically motivated restriction directly into the axiomatic system, you can place a restriction on a rule of  $\text{KT4M} \oplus (\text{LP})$ . A suitable restriction would be the metarule that you cannot apply the rule of necessitation N to anything you derive after (LP) has introduced a Gestalt shift. In  $\text{KT4M}_T$ , N bestows truth-regardless on any formula that has been derived in the system. For the above reasons, once we have a Gestalt shift, such a move to truth-regardless is no longer justifiable. Thus Gestalt shifts are accommodated by putting a restriction on a rule of KT4M for the system  $\text{KT4M} \oplus (\text{LP})$ . In the above argument, this restriction on N bans the step from 5 to 6. As a result of supplementing KT4M with (LP) and an (LP)-based restriction on N, you obtain a logic that is extremely close both to the Kripke-Feferman system (KF) and to the formal theory Herzberger outlines as emerging from his Naive Semantics (Feferman 1984, 1991; Herzberger 1982a, p. 493). One crucial difference in the suggested theory is axiom M.

and  $T\neg L$  as expressions that capture the difference in pragmatic aspect (in what is salient), then we can add the following from a logical perspective: we would expect that if you add  $L$  and  $T\neg L$  as hypotheses in derivations in the constrained C-KT4M, then  $L$  and  $T\neg L$  have exactly the same inferential power. That is, that you can derive no more or less from one than from the other. And this is indeed the case. In KT4M, from  $L$  as well as from  $T\neg L$  you can derive exactly the same things, which is precisely nothing that you could not derive with any other sentence of  $\mathcal{L}_T$  in place of  $L$  or of  $T\neg L$ . This can be easily shown.<sup>33</sup> Thus C-KT4M both removes the inconsistencies and retains the relevant information about the liar sentence. One main purpose of sections 4.1 to 4.7 was to bring this out.

It is not the liar sentence ( $L$ ), but those two ways of understanding it, and that is a pragmatic fact, that license the derivation of incompatible conclusions in the hypothetical deductions or *arguments 1 to 4*. This fact is what makes the liar sentence paradoxical. In (12) from section 4.10 and in (LP) (when read as expressing a nonpragmatic property) we have a confusion of a pragmatic relation with a logical or semantic relation. So, truth is not an inconsistent notion (see above). Semantic paradoxes like the liar sentence make it appear inconsistent. They make it *appear* inconsistent, because a pragmatic feature (the bistability of liar sentences) is wrongly taken to be a nonpragmatic relation that can be accurately expressed in a material biconditional. ( $L$ ) is disqualified as a semantically evaluable sentence, just as an ambiguous sentence is disqualified as an atomic sentence of classical logic.

#### 4.11. The Strengthened Liar and Revenge

It is not uncommon for discussions of the Liar, or theories of truth, to ditch the simple Liar ( $L$ ) and other semantic-value-self-ascribing sentences that self-ascribe falsehood for a sentence like

(13) This sentence is untrue.

---

<sup>33</sup> Thus you can derive e.g.  $L \vee \neg L$ ,  $T\neg L \vee \neg T\neg L$ . By contrast, from the hypothesis  $L$  alone you can derive nothing, given  $\mathbf{FL}$  (or  $\mathbf{FL}_R$ ) and axiom M. On  $T\neg L$  you can apply any axioms or rules of C-KT4M. Since we cannot prove the semantic value of  $T\neg L$ , due to  $\mathbf{FL}$ , MIRROR, and axiom M, we can detach no consequent of any conditional in which  $T\neg L$  is the antecedent (or from a set of sentences  $\Delta$  that includes it) and for the same reason we cannot apply N. On the other hand, from the hypothesis of an assessment-insensitive  $p$  you can derive  $p$  ( $Tp$ ,  $p$ ), and from the hypothesis of  $T\neg p$  you can derive  $\neg p$ .



One rationale behind this is the following. Often, the paradoxicality of the simple Liar (L) is removed by the introduction of a third semantic status, only to return with a vengeance, armored in the garb of the strengthened liar. (13) preempts the revenge in such simple(ton) solutions. So, why not start where the real problem lies? The reason *my* approach starts with the simple Liar is that this made it easier to bring out what I suggest is the structure of the notion of truth.

Even so, (102) since, in *some* sense, the proposed theory offers the possibility of three semantic statuses for sentences, I should say something about the strengthened liar. The traditional way of introducing revenge would be by means of a sentence that is a disjunction of the two values that are not the designated value of truth.<sup>34</sup> So let us do that for the representation of truth as truth-regardless:

$$(L_R') \quad T \neg L_R' \vee \mathbf{F}L_R'$$

In KT4M, (L<sub>R</sub>') is logically equivalent to<sup>35</sup>

$$(L_R) \quad \neg \mathbf{T}L_R$$

In natural language, (L<sub>R</sub>) can be expressed as ‘This sentence is not true-regardless’. (L<sub>R</sub>) is the expected candidate for the modal version of the strengthened-liar sentence. It is easy to check that, like the simple Liar, it is bistable, context sensitive, and assessment sensitive. Like ‘L’ and ‘ $\neg \mathbf{T}L$ ’, ‘L<sub>R</sub>’ and ‘ $\neg \mathbf{T}L_R$ ’ could be taken to indicate the Gestalt in which the sentence is understood. There is then also a strengthened-liar property, that L<sub>R</sub> *is*  $\neg \mathbf{T}L_R$ . Like the simple liar property, this strengthened-liar property is taken to be not a semantic but a pragmatic feature which allows a *prima facie* pragmatically licit transition from one sentence to the other.

---

<sup>34</sup> Revenge arguments tend to be tailor-made to each liar solution. (See, e.g., Beall 2007 for a whole spectrum of vengeance.) Here I only show that the present theory accommodates the standard way of introducing revenge.

<sup>35</sup> The proof is trivial and is left to the reader. (Hint: use De Morgan, contraposition,  $\wedge$ -elimination, def.  $\leftrightarrow$ , def.  $\mathbf{F}$ .)

Again, parallel to the case of the simple Liar, nothing prevents you from syntactically expressing this pragmatic strengthened-liar property as a biconditional in a language  $\mathcal{L}$ :

$$(L_R P) \quad L_R \leftrightarrow \neg TL_R$$

with the two conditionals of  $(L_R P)$ , denoted as  $(L_R P \rightarrow)$  and  $(L_R P \leftarrow)$ , taken to indicate Gestalt shifts. Adding the pragmatic property  $(L_R P)$  as a theorem to KT4M, you can then express a revenge paradox, for example

1.	$\neg TL_R \rightarrow L_R$	$(L_R P \leftarrow)$ (stated)
2.	$TL_R \rightarrow L_R$	axiom T
3.	$L_R$	2, PC
4.	$\neg TL_R$	3, $(L_R P \rightarrow)$ (used)
5.	$TL_R$	3, Necessitation
6.	$TL_R \wedge \neg TL_R$	4, 5, $\wedge$ -introduction

In order to remove the resulting inconsistencies, you can further supplement the system  $KT4M \oplus (L_R P)$  with a revision-theoretical rule (RstrLiar), adjusting the rule (RLiar) or the rule (RLiar') in a suitable way.

Again, (103) the addition of  $(L_R P)$  and a revision-theoretical rule to KT4M would be sorely misguided. This is so, since the pragmatic liar property  $(L_R P)$ , too, is already accounted for in KT4M via the viewpoint-interpretation of the possible-worlds semantics. (13) is context sensitive and viewpoint sensitive, just like the simple Liar. Those two features are represented in a consistent way in  $KT4M_T$ .<sup>36</sup> The only logical differences between (L) and (13) are those that hold between the schemata  $T\neg A$  and  $\neg TA$  in KT4M. The strengthened-liar sentence  $(L_R)$  introduces no further possible semantic statuses beyond true-regardless, false-regardless and true-depending. Accordingly there is no iteration or proliferation of revenge sentences.

---

<sup>36</sup> Arguments analogous to *Arguments 1* to *4* can be constructed, with the context fixed, in the  $KT4M_T$  semantics, by means of a circumstance-conjunct. With either circumstance-conjunct, there are viewpoints at which it can be inferred that  $(L_R)$  is not true, and others at which it can be inferred that  $(L_R)$  is not false.

To sum up sections 4.10 and 4.11: The present theory acknowledges that there is a pragmatic liar property *of self-reference*, in the sense that—with the context fixed—the sentence which ascribes falsehood (untruth) to itself *is* the sentence which is denoted by the subject expression of the sentence which ascribes falsehood (untruth) to itself—just from a different viewpoint. Liar sentences, whether simple or strengthened, do only apparently lead to contradiction, via a confusion of a pragmatic characteristic with a nonpragmatic one. The pragmatic feature disqualifies the sentences from being semantically evaluable.

The reasoning in this and the previous section can with some modification be applied to the standard argument for an inconsistency in the truth predicate  $Tr$  that purports to produce, via Diagonal Lemma and Incompleteness Theorem, and relative to some language  $\mathcal{L}$ , the contradiction

$$(14) \quad Tr(\ulcorner \neg Tr \urcorner) \leftrightarrow \neg Tr(\ulcorner \neg Tr \urcorner).$$

But here is not the place to do this.<sup>37</sup> Equally, truth represented by  $KT4M_T$  can accommodate multiple-sentence liar paradoxes, propositional versions of Kripke’s Nixon/Dean example, Curry Sentences, and open sentences such as the No-no-Paradox. Again, for reasons of space, these cases cannot be discussed. The procedure is always based on bistability and is similar to those for (T), (L), and (L<sub>R</sub>), or combinations of these.

---

<sup>37</sup> As is obvious from this chapter, I do not believe that the so-called naïve theory of truth is our theory of truth. So, with regard to diagonalization, it would seem entirely reasonable to remove Liar sentences from the relevant language (e.g., PA with truth predicate) for the reasons (i) that it cannot be established that such sentences are semantically evaluable and (ii) that Tarski biconditionals do not extend to Liar sentences (see section 14). However, it seems more satisfying to show that liar paradoxes that are based on the diagonal lemma to prove the inconsistency of some  $\lambda$ , too, make use of the bistability of  $\lambda$ , and that  $\lambda$  is no less assessment-sensitive than our (L) and (L<sub>R</sub>). This is possible. The argument leaves the strengthened and extended diagonal lemma intact. The problem rather lies in the way the Tarski biconditional for the liar sentence is employed in the paradox. To those who at this point bring up Montague’s Theorem and the paradox of the knower, I point out that they are changing the subject. The predicate introduced by diagonalization in the proof of Montague’s Theorem is different from the truth predicate in the sense that we expect fewer sentences to be derivable. Perhaps what is knowable or informally provable is true, but the reverse does not hold. Whereas a modal logic for truth should be normal and complete, the same is not obvious for a knowability or informal-provability operator. (See also footnote 2.)

## 4.12. KT4M<sub>T</sub> as the Logic of Semantic Modalities: Bivalence and LEM Come Apart

The reflections on the Gestalt shifts, context sensitivity, and assessment sensitivity of the Liar do not just provide a solution to the liar paradox. They also provide new insight into the notion of truth in natural language. Sections 4.12 to 4.14 examine more closely how KT4M is related to truth.

By (104) analogy with the logics of logical modalities and metaphysical modalities, KT4M can be called *the logic of semantic modalities* (cf. section 4.13 below). This makes truth semantic necessity and bivalence semantic noncontingency. Just as a logic of metaphysical modalities does not produce individual metaphysical necessities as theorems, so a logic of semantic modalities does not produce individual truths as theorems. Rather, it defines the *logical structure* of truth and bivalence.

Nonetheless, the logic of semantic modalities KT4M can be employed to gain further insight about semantic paradoxes, and truth, when we make the distinction between assessment-sensitive and assessment insensitive sentences explicit.

**Bivalence and excluded middle.** In Section 4.9 we saw that C-KT4M validates that *all semantically non-paradoxical atomic sentences are bivalent* (Result 1), that *no atomic assessment-sensitive sentence  $p$  or its negation is valid in C-KT4M* (Result 2), that *of no assessment-sensitive sentence  $p$  is any  $T^m F^n p$  or any  $T^m F^n \neg p$  (with  $m \geq 0$  and  $n \geq 1$ ) valid in C-KT4M* (Result 3), and moreover that *for Liar sentences, bivalence can neither be validated nor invalidated* (Result 4). Hence, given the assumption of classical logic, *the principle of bivalence and the law of excluded middle  $A \vee \neg A$  (LEM) come apart* (Result 5). For any sentence  $A$ , liar sentences included,  $A \vee \neg A$  is valid in KT4M. On the other hand, for liar sentences the modally expressed principle of bivalence,  $TA \vee T\neg A$ , is not valid in KT4M.<sup>38</sup> (A similar result holds for  $TA \leftrightarrow A$ , see Section 4.14.)

**Truth and falsehood.** In C-KT4M, neither  $Tp$  nor  $T\neg p$  is valid of any atomic sentence  $p$ . This is as it should be. The truth-regardless (truth *tout court*) and falsehood-regardless (falsehood *tout court*) of atomic sentences are not structural properties of the notion of truth. Even if we know that a sentence is nonparadoxical, we cannot infer its truth or its falsehood from that. Accordingly, even in the C- KT4M, truth and falsehood can only be defined relative

---

<sup>38</sup> The proposed theory shares the feature that the principle of bivalence and LEM come apart with some other theories of truth, such as supervaluationist ones (e.g., van Fraassen 1968 and 1970; also Fine 1975).

to a model: *a sentence A is true-regardless (TA) in a model M precisely if it is bivalent and is valid-in-M* (Result 6). From the definition of true-regardless, together with the notions of negation and circumstance-conjuncts, it results that *a sentence A is false-regardless (T¬A) in a model M precisely if it is bivalent and its negation is valid-in-M* (Result 7). This is in line with section 4.5 and so is also as it should be.

### 4.13. KT4M as the Logic of Semantic Modalities and the Semi-determinability of Truth

As I said above, by analogy with the logics of logical modalities and of metaphysical modalities, KT4M can be called *the logic of semantic modalities*. This makes truth semantic necessity and bivalence semantic noncontingency. We can also say that KT4M defines the logical structure of truth.

Assume (104) that it is the mark of a deflationary theory of truth that truth does not denote a real property of sentences (or propositions); or that there is no such thing as the nature of truth; or that saying that ‘snow is white’ is equivalent to ‘it is true that snow is white’ or to “‘snow is white’ is true’ exhausts what one can meaningfully say about the truth of ‘snow is white’—and so for all sentences. Then we would expect deflationists to hold that TRIV is the modal system that captures the structure of truth and provides the logic of semantic modalities.<sup>39</sup> This would be a modal way of expressing that the notion of truth has no content. The theory I propose is not a deflationary theory. It endows the notion of truth with content (as presumably would any modal logic stronger than TRIV). This content is entirely structural, but it is content nonetheless. Let me expound this content of truth in two steps.

*First*, the content of truth can be said to be bivalence. As was mentioned in section 4.7, the negation of A’s being true-depending, or  $\mathbf{FA}$ , is the bivalence of A. In modal terms,  $\neg\mathbf{FA}$  expresses that either it is true that A or it is false that A. And this is precisely that A is bivalent. Making use of the interdefinability of  $\mathbf{T}$  and  $\mathbf{F}$ , one obtains the following account of truth:

$$(15) \quad \mathbf{TA} \leftrightarrow \neg\mathbf{FA} \wedge A$$

---

<sup>39</sup> TRIV is any normal modal system that contains axiom K and in which  $\Box A \leftrightarrow A$  is a theorem.

Since KT4M can be defined with the basic operator  $\mathbf{F}$  instead of  $\mathbf{T}$  we can also express this as a definition of truth:

$$(16) \quad \mathbf{TA} =_{def} \neg \mathbf{FA} \wedge A \quad ^{40}$$

We can then say that every sentence  $\mathbf{TA}$  has as its content bivalence, but not necessarily every sentence  $A$  does.<sup>48</sup>

*Second*, the logical structure of bivalence qua content of truth is defined by the modal logic KT4M. What kind of structure does KT4M equip bivalence (and truth) with? The simplest way to understand this structure is as representing a kind of semideterminability, or more precisely proper semideterminability, with respect to bivalence (and truth) over classical logic as applied to the atomic sentences of the natural language fragment  $\mathcal{L}_T$  with a truth-predicate, where these atomic sentences include liar and truth-teller sentences. By proper semideterminability I mean the following:

A class of questions is *properly semideterminable* if and only if there is a procedure that comes to a halt and says yes if the answer is positive, but there is no procedure that comes to a halt and says no if the answer is negative.<sup>41</sup>

Rather than setting out a full mechanism that results in such proper semideterminability for bivalence and truth, I offer an informal description that relies on the partially interpreted possible-worlds semantics for C-KT4M from section 4.9.

For (106) bivalence, the class of questions at issue asks of every sentence of  $\mathcal{L}_T$  whether it is bivalent. The procedure uses *the partially interpreted semantics* for C-KT4M given above. This assigns in every model to every atomic sentence  $p$  a circumstance-conjunct, and to every viewpoint in that model an assessment argument that assigns at that viewpoint a semantic value to  $p$  and then restricts consideration to the class of models in which each

---

<sup>40</sup> Mutatis mutandis the same holds for falsehood. For a sentence ‘S’ to be false means that ‘S’ is bivalent and that it’s not the case that S.

$$\mathbf{T}\neg A =_{def} \neg \mathbf{FA} \wedge \neg A$$

This follows from the modal definition of falsehood.

<sup>41</sup> Herzberger (1982, p. 496), introduces a relation very similar to (15) in his logic sketch in schema ii. He does not consider the question of the content of the notion of truth.

<sup>42</sup> Some will recognize this as the informal definition of non-decidable semidecidability or recursive enumerability. Here I am interested solely in this informal non-mathematical definition, hence the different term.

sentence letter in the interpretation is constantly through the models either given an assessment-sensitive or an assessment-insensitive sentence. If, on this semantics for the relevant subclass of models, for an  $A$ , the expression  $TA \vee \neg TA$  can be shown to be valid, then  $A$  is bivalent, and the answer to the semi-determinability question is yes. For all assessment-insensitive sentences this can be shown (Section 4.9). There is no corresponding procedure that rules out bivalence for sentences that are not bivalent, that is, for the argument-based assessment-sensitive sentences: since if  $FA$ , then  $F^n A$  for any  $n$ , for no  $n$  can  $F^n A$  come out as valid. So there is no procedure that results in a ‘no’ for these cases (sections 4.9 and 4.12).

The case for the semideterminability of truth is similar. For truth, the class of questions at issue asks of every sentence of  $\mathcal{L}_T$  whether it is true. Since the semantic value of  $p$  depends on the circumstance-conjunct for  $p$  (to ascertain whether an atomic sentence  $p$  is true, we need to know whether things are as  $p$  says, and that is, we need to know the circumstance-conjunct for  $p$ ), here the restriction is tightened further to a class of models of KT4M in which each atomic sentence of  $\mathcal{L}_T$  is assigned the same circumstance-conjunct across models. If  $A$  is true – i.e. if we have  $TA$  – in every model of this set, then the semideterminability question is answered with yes. Since falsehood is defined as  $T\neg A$ , indirectly it is also covered. Thus for assessment-insensitive sentences this procedure provides an answer. But again, there is no corresponding procedure that shows of sentences that neither are true nor have true contradictories in every model of this class that they are not true in this class of models. And that is, there is no procedure that results in a ‘no’ for such sentences. These sentences will be the assessment-sensitive ones. (This again reflects the deliberation from section 4.9)

Reflections on the Liar Paradox thus result in a theory of truth that offers an explanation how the ‘nature’ of truth, or the content of the notion of truth, is a structural property, i.e. bivalence. Bivalence itself is defined with the help of the modal system KT4M, and is semideterminable – as, consequently, is truth. Deflationary theories of truth turn out to be inadequate to capture truth.

#### 4.14. The T-Schema, Convention T, and Coherence

How does the KT4M account of truth fare with regard to the T-Schema, which can be informally expressed as ‘‘ $\ulcorner S \urcorner$  is true if, and only if,  $S$ ’ where ‘ $\ulcorner S \urcorner$ ’ is the name of the sentence abbreviated as  $S$ ? And how does it fare with regard to Tarski’s (107) Convention T, or

material adequacy condition, that any acceptable theory of truth needs to entail a sentence of the form of the T-Schema for every sentence  $S$  of the language at issue (where ‘ $\ulcorner S \urcorner$ ’ is the name of the sentence  $S$  in the metalanguage, with  $S$  being a translation of the corresponding sentence in the object language)? Since I express truth by means of an operator instead of a predicate, results can only be analogs to this requirement.

KT4M directly provides the analogue to the left-to-right conditional of the T-Schema, since it contains axiom T. Axiom T holds for all sentences, including liar sentences. If it is true that  $\ulcorner L \urcorner$  is false, then  $\ulcorner L \urcorner$  is false. Whether  $\ulcorner L \urcorner$  is false may of course be viewpoint-dependent. So may be whether it is true that  $\ulcorner L \urcorner$  is false.

The analogue to the right-to-left conditional cannot be derived in KT4M. For a liar sentence  $\ulcorner L \urcorner$ , one cannot rule out that  $L$  but not  $TL$ . So, one cannot prove that the T-schema does not hold for liar sentences. One cannot prove that the T-schema *does* hold for liar sentence either. Since, as intended, other than the PC tautologies, one cannot demonstrate anything for liar sentences within C-KT4M, the proposed theory entails that *one cannot rule out that (the analogue to) the T-schema holds universally*.<sup>43</sup>

So, as in Herzberger’s revision theory (1982a, p. 493), in the theory presented here the T-schema cannot be shown to be valid. Unlike Herzberger’s, the theory does not entail that the T-schema is invalid. My view is that *theories that satisfy Convention T get things wrong in the sense that they provide an idealization rather than a representation of truth in natural language*. By contrast, the proposed theory gets things just right, because – *given that Liar sentences are part of natural language* – it is neither possible to show that the T-schema holds for all natural-language sentences, nor to show for any sentence that it does not hold for it. Among other things, the theory is a representation of this fact. Where the T-Conventioners provide an *idealization* of truth by removing some of its structural content, the present theory aims to provide a *representation* of truth, leaving that somewhat perplexing structural content in place. (How can I do this when I take liar and truth-teller sentences as atomic? By offering an axiomatic theory that describes the logical structure of their predicates.)

The proposed theory thus stands to Tarski’s as follows. Both theories entail that it is not possible to show that Convention T holds for truth in natural language. Where Tarski

---

<sup>43</sup> Thus, unlike Liggins (2014), who suggests that the semantic-ascent half of the T-schema should be renounced, all I say is that judgment needs to be suspended regarding its validity.



states that Convention T does not hold for truth in natural language, the proposed theory suggests that this question is undecidable and commends suspension of judgement on the issue. As for coherence, I disagree with Tarski’s view that truth in natural language is incoherent and have given a formal (untyped) representation of truth in the natural-language fragment  $\mathcal{L}_T$  that shows that it is coherent. My view is that what matters is not Convention T, but rather (i) that a coherent representation of truth can be given and (ii) that the T-schema fails for no other sentences than those nobody intentionally asserts (and even for those it fails only in the direction of semantic ascent). Point (i) has been shown. Point (ii) comes next.

***Liar*** (108) ***sentences can be described and picked out independently of the representation of truth by KT4M.*** They are self-referring sentences that predicatively self-ascribe a semantic value that is incompatible with truth. It is a characteristic feature of all such sentences that nobody intentionally asserts them except, perhaps, hypothetically when contemplating or discussing liar sentences. (Why this is so is not my present concern.) Thus it is questionable whether there is *any point* in requiring a truth theory to be capable of showing that liar sentences satisfy the semantic-ascent half of the T-schema.

In the C-KT4M (section 4.9), because of their viewpoint sensitivity, liar sentences are *represented as (not shown to be!)* the sentences for which  $\mathbf{TA}$  holds. If one removes all such sentences from consideration, one obtains the fragment of sentences for which  $\mathbf{TA} \vee \mathbf{T}\neg\mathbf{A}$  is valid. In this fragment of  $\text{KT4M}_T$ , given the definition of  $\mathbf{F}$ ,  $\neg\mathbf{TA}$  and  $\mathbf{T}\neg\mathbf{A}$  mutually entail each other, so that we have  $\neg\mathbf{TA} \leftrightarrow \mathbf{T}\neg\mathbf{A}$ . The semantic modality  $\mathbf{T}$  then collapses and we obtain  $\mathbf{A} \leftrightarrow \mathbf{TA}$ .<sup>44</sup> In this trivial fragment, or trivialization, of  $\text{KT4M}$ , one can understand the  $\mathbf{T}$ -operator as ‘truth for use’, since arguably there is no use for liar sentences outside of the discussion of the liar paradox—and in such situations it is apt to employ  $\text{KT4M}$  as a whole.

---

44

- |    |  |   |
|----|--|---|
| 1. | $\neg\mathbf{TA} \leftrightarrow \mathbf{T}\neg\mathbf{A}$ |   |
| 2. | $\mathbf{T}\neg\mathbf{A} \rightarrow \neg\mathbf{A}$      | 1, axiom T  |
| 3. | $\neg\mathbf{TA} \rightarrow \neg\mathbf{A}$               | 1, 2, definition of $\leftrightarrow$                             |
| 4. | $\mathbf{A} \rightarrow \mathbf{TA}$                       | 3, contraposition   |
| 5. | $\mathbf{TA} \rightarrow \mathbf{A}$                       | axiom T   |
| 6. | $\mathbf{A} \leftrightarrow \mathbf{TA}$                   | 4, 5, $\wedge$ -introduction, definition of ‘ $\leftrightarrow$ ’ |

Having  $\mathbf{A} \leftrightarrow \mathbf{TA}$  as a theorem in the trivial modal fragment, or trivialization, of  $\text{KT4M}_T$  is not strictly the same as having the T-schema for the relevant sentences  $\mathbf{A}$ . For this one needs to add that, for any sentence  $\mathbf{A}$  we have ‘ $\mathbf{TA}$  if and only if’  $\boxed{\mathbf{A}}$  ‘is true’.

## Acknowledgments

I thank Walter Dean, Øystein Linnebo, Stephen Menn, Beau Madison Mount, Ian Phillips, Ian Rumfitt, Josh Schechter, and Bruno Whittle for helpful comments on parts of what has become this chapter, Simona Aimar for reading an early draft and for asking penetrating questions, Bradley Armour-Garb for notes on two drafts and for his saintly patience and encouragement as editor, and in particular Kit Fine, Volker Halbach and whose avid discussions of earlier versions of what has become this chapter have benefited it enormously. I do not thank the anonymous referee from Oxford University Press.

## References

- Barwise, Jon, and Etchemendy, John. 1987. *The Liar: An Essay on Truth and Circularity*. New York: Oxford University Press.
- Beall, J. C. 2007. *The Revenge of the Liar*. Oxford: Oxford University Press.
- Belnap, Noel, and Anil Gupta. 1993. *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Billon, Alexandre. 2011. “My Own Truth: Pathologies of Self-Reference and Relative Truth.” In Shahid Rahman, Giuseppe Primiero and Mathieu Marion, eds. *The Realism-Antirealism Debate in the Age of Alternative Logics*, 25-45. New York: Springer 2011, Coll. Logic, Epistemology, and the Unity of Science (LEUS), 23.
- Bobzien, Susanne. 2015. “Columnar Higher-Order Vagueness or Vagueness *Is* Higher-Order Vagueness.” *Aristotelian Society Supplementary Volume* 89: 61–87.
- . 2010. “Higher-Order Vagueness, Radical Unclarity, and Absolute Agnosticism.” *Philosophers’ Imprint* 10 (10): 1–30.
- Feferman, Solomon. 1984. “Toward Useful Type-Free Theories, I.” *Journal of Symbolic Logic* 49: 75–111.
- . 1991. “Reflecting on Incompleteness.” *Journal of Symbolic Logic* 56: 1–49.
- Fine, Kit. 1975. “Vagueness, Truth and Logic.” *Synthese* 30: 265–300. Reprinted with corrections in Rosanna Keefe and Peter Smith, eds., *Vagueness: A Reader* (Cambridge, MA: MIT Press, 1996), 119–50.
- Glanzberg, Michael. 2004. “Truth, Reflection, and Hierarchies.” *Synthese* 142: 289–315.
- Halbach, Volker. 2003. “Modalized Disquotationalism.” In V. Halbach and L. Horsten, eds., *Principle of Truth*, 75–102. London: Ontos Verlag.

- Halbach, Volker, and Leon Horsten, eds. 2003. *Principle of Truth*. London: Ontos Verlag.
- Halbach, Volker, Hannes Leitgeb, and Phillip Welch. 2005. “Possible Worlds Semantics for Predicates.” In R. Kahle, ed., *Intensionality: Lecture Notes in Logic*, 20–41. Wellesley, MA: A K Peters.
- Herzberger, Hans G. 1970. ‘Paradoxes of Grounding in Semantics’. *Journal of Philosophy* 17: 145-167.
- Herzberger, Hans G. 1982a. “Naive Semantics and the Liar Paradox.” *Journal of Philosophy* 79: 479–497.
- . 1982b. “Notes on Naive Semantics.” *Journal of Philosophical Logic* 11: 61–102.
- . 2011. *The Tarskian Turn: Deflationism and Axiomatic Truth*. Cambridge, MA: MIT Press.
- Hughes, (113) G. E., and Max J. Cresswell. 1996. *A New Introduction to Modal Logic*. London: Routledge.
- Kripke, Saul. 1975. “Outline of a Theory of Truth.” *Journal of Philosophy* 72: 690–716.
- Leitgeb, Hannes. 2003. “Metaworlds: A Possible-Worlds Semantics for Truth.” In V. Halbach and L. Horsten, eds., *Principle of Truth*, 129–152. London: Ontos Verlag.
- Liggins, David. 2014. “Constructive Methodological Deflationism, Dialetheism, and the Liar.” *Analysis* 74 (4): 566–574.
- MacFarlane John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.
- McGee, Vann. 1991. *Truth, Vagueness, and Paradox*. Indianapolis: Hackett.
- Parsons, Charles. 1974. “The Liar Paradox.” *Journal of Philosophical Logic* 3: 381–412.
- Ryle, Gilbert. 1951. “Heterology.” *Analysis* 11: 61–69.
- Summerfield, C., and T. Egner. 2009. “Expectation (and Attention) in Visual Cognition.” *Trends in Cognitive Science* 13 (9): 403–409.
- Tarski, Alfred. 1935. “Der Wahrheitsbegriff in den formalisierten Sprachen.” *Studia Philosophica* 1: 261–405.
- van Fraassen, B. 1968. “Presupposition, Implication, and Self-Reference.” *Journal of Philosophy* 65: 136–152.
- van Fraassen, B. 1970. “Truth and paradoxical consequence”, in Martin, Robert L. (ed.), *The Paradox of the Liar*, Atascadero: Ridgeview. 1970:13–23.