Sponsored since 2011 by the
Italian Society for Analytic Philosophy

ISSN 2037-4445
http://www.rifanalitica.it

A New Model for the Liar

Luca Castaldo[*]

Abstract. A new model for $\mathscr{L}^t_{pa}$ (the language of arithmetic enhanced by the unary truth predicate T) is presented, which extends Kripke's minimal fixed point. The latter, it will be argued, does not adequately model the truth predicate, since no difference between Liars and Truth-tellers can be made. The new model, which contains an extension of Kripke's interpretation of T along with a new 4-valued logic, overcomes this inadequacy. The gist of my proposal is that 'paradoxical' ought to be treated as a truth value: Liar sentences, unlike Truth-teller sentences, do not simply *lack* a truth value. They do posses one: they are *paradoxical*.

Keywords. Formal theories of truth, Liar paradox, 4-valued logic, Kripke.

[*] I wish to thank Professor Karl-Georg Niebergall for valuable comments and suggestions.

Author. Luca Castaldo, castaldx@gmail.com.
Received. 7 September 2016. **Accepted**. 29 October 2016.

Introduction

It seems that

> **P** Every sentence is *either* true *or* untrue[1].

But, if so, what about the Liar sentence below?

> **$** The sentence marked with a dollar is untrue.

$ contradicts **P**, for it is true *if, and only if,* it is untrue. To (dis)solve the problem, Kripke (1975) proposes to reject **P** and, exploiting the 3-valued logic called *Strong Kleene*, constructs a partial model for the truth predicate, where sentences like **$** are 'undefined', i.e. they *lack* a truth value.

Within Kripke's model, however, also the so-called Truth-teller

> € The sentence marked with a euro is true.

lacks a truth value. Yet, **$** and **€** are, admittedly, very different: the latter can *consistently* be declared true or untrue; the former cannot. An adequate model for the truth predicate ought to account for their diversity.

The purpose of this paper is to put forward a new response to the Liar paradox, which extends and improves the work done by Saul Kripke in his seminal *Outline of a Theory of Truth*.

The plan is as follows: after technical preliminaries in § 1 (including the construction of the formal Liar sentence), I go on in § 2 to present a new model for the truth predicate along with a new 4-valued logic, thereby proposing the new response to the Liar paradox. The final section 3 examines the properties of the model, proving what I shall call 'metalinguistic T-Schema'.

A last remark before I begin: In what follows I assume the reader is familiar with (i) *Peano arithmetic*, (ii) the *arithmetization of syntax*, and (iii) Kripke's *Outline of a Theory of Truth*[2].

[1] The *either ... or* is to be read here as an exclusive disjunction.

[2] There is an extensive literature on Kripke's *Outline*. A more philosophical and informal introduction is offered by Burgess, (2011). For more information on the mathematical aspects of Kripke's construction see, for example, Fitting, (1986) and McGee, (1991, §§4-5). The axiomatic theory known as Kripke-Feferman (**KF**) was first given by Reinhardt, (1986) and Feferman, (1991). Feferman, (1991) also determines its proof-theoretic strength. Cantini, (1989) gives a more direct proof-theoretic analysis of **KF** and some of its subsystems. In **KF**, the *partial* notion of truth advanced by Kripke is axiomatised in *classical* logic. Therefore, outer logic (what is provable) and inner logic (what is provably true) of that system differs substantially. Halbach and Horsten, (2006) (see also Horsten, 2011, §9.5) have proposed an interesting axiomatisation in partial logic, creating a system, called "partial Kripke-Feferman" (**PKF**), within which the two logics coincide. In that system, gaps but no gluts are admitted. Halbach, (2014, §16) proposes a system that admits both. For critical discussions of Kripke's position see, among others, Gupta, (1982, pp. 30-37) and Field, (2008, §3).

1 The Formalised Liar

1.1 Technical Preliminaries

The object language of this work will be the language of Peano arithmetic (**PA**) extended by the unary truth predicate T. I shall call the language of **PA**, without T, $\mathscr{L}_{pa}$; the extended language will be called $\mathscr{L}^t_{pa}$[3]. As "official" logical vocabulary, I shall use the existential quantifier $\exists$, the negation and disjunction symbols $\neg$, $\vee$, and the identity symbol $\doteq$. As usual, however, abbreviations will be used. A standard *Gödel numbering* of $\mathscr{L}^t_{pa}$-expressions will be assumed throughout the work, without going into details[4]. The Gödel number (or code) of a formula φ is $gn(\varphi)$, and $\ulcorner\varphi\urcorner$ is the numeral of $gn(\varphi)$. I shall distinguish between natural numbers and $\mathscr{L}^t_{pa}$-numerals exploiting boldfaced characters: the natural numbers are written "$0, 1, 2, \ldots, n$" (not boldfaced) and the $\mathscr{L}^t_{pa}$-numerals "$\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n}$" (boldfaced), where "$\mathbf{1}, \mathbf{2}, \mathbf{3} \ldots$" abbreviates "$\mathbf{0}', \mathbf{0}'', \mathbf{0}''' \ldots$". Formulae with one free variable are indicated by $\varphi(v_i)$; $\varphi(t)$ denotes $\varphi[t/v_i]$, i.e. the result of substituting t for v_i in φ. I write $\varphi \equiv \psi$ to indicate that φ and ψ are names of the same formula.

$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle$ is Kripke's minimal fixed point (henceforth: MFP), and '$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle \models_{sk} \varphi$' means that φ is true in MFP, according to the Strong Kleene. Furthermore, I shall make use of the following metalinguistic symbols:

- $\neg\!\!\!\neg$ for "non …".
- $\vee\!\!\!\vee$ for "… or …".
- $\wedge\!\!\!\wedge$ for "… and …".
- $\Rightarrow$ for "if …, then …".
- $\Leftrightarrow$ for "… if, and only if, …".
- $\exists\!\!\!\exists$ for "there is …".
- $\forall\!\!\!\forall$ for "for all …".

1.2 $\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$

The *Diagonal Lemma*[5] is, as McGee, (1991, p. 24) put it, "a cornerstone of modern logic". He even adds that "most of the results of [*Truth, Vagueness, and Paradox*] can be regarded as corollaries to this basic result". In this section I shall exploit the typical diagonal construction, in order to obtain the formalised liar antinomy.

[3]Notice that we are just extending the language of **PA**, not the theory, i.e. we are not adding axioms for T, creating a new theory, say $\mathbf{PA^T}$. In addition, we can impose a restriction on the induction schema to $\mathscr{L}_{pa}$-formulae, i.e., an instance of

$$(\varphi(\mathbf{0}) \wedge \forall v_i(\varphi(v_i) \to \varphi(v_i'))) \to \forall v_i(\varphi(v_i))$$

is an axiom, only if T does not occur in φ.

[4]See, for instance, Boolos, Burgess, and Jeffrey, (2007) and Smith, (2013).

[5]Or *Fixed Point Lemma*, or *Self-Referential Lemma*.

Before I begin, the concept of *diagonalization* of a formula must be introduced:

> The *diagonalization* of φ is the expression $\exists v_0(v_0 \doteq \ulcorner\varphi\urcorner \wedge \varphi)$.

Even if this notion makes sense for arbitrary expressions, it is of most interest in the case of a formula $\varphi(v_0)$ with just one variable v_0 free. Since an expression of the form $\varphi(t)$ is equivalent to $\exists v_0(v_0 \doteq t \wedge \varphi(v_0))$, the diagonalization of $\varphi(v_0)$ is equivalent to $\varphi(\ulcorner\varphi\urcorner)$. That is: the diagonalization of a formula $\varphi(v_0)$ is true (in the standard interpretation) if, and only if, it is satisfied by its own code.

There is also a recursive function $diag$ that, when applied to the Gödel number of a formula, yields the Gödel number of its diagonalization. That is to say: if the code of a formula φ is n and the code of its diagonalization is m, then $diag(n) = m$. A more formal definition is:

$$diag(n) = gn[\exists v_0(v_0 \doteq\ ] \star num[n] \star gn[\ \wedge\ ] \star n \star gn[)],$$

where $\star$ and num represent, respectively, the concatenation and the numeral functions, both recursive[6].

Lemma 1.1. (The Formalised Liar) There is a $\mathscr{L}^t_{pa}$-sentence λ, such that

$$\mathbf{PA} \vdash \lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$$

Proof. Since **PA** represents every primitive recursive function, $diag$ is representable in **PA**. Let $\mathbf{Diag}(v_0, v_1)$ be a formula representing $diag$, so that for any a and b, if $diag(a) = b$, then

$$\mathbf{PA} \vdash \forall v_1(\mathbf{Diag}(\mathbf{a}, v_1) \leftrightarrow v_1 \doteq \mathbf{b}) \tag{1}$$

Diag is a complex $\mathscr{L}_{pa}$-formula, *not* containing the new predicate T.

Let now $\beta(v_0)$ be the formula

$$\exists v_1(\mathbf{Diag}(v_0, v_1) \wedge \neg T(v_1)) \tag{$\beta(v_0)$}$$

Intuitively, $\beta(v_0)$ says that the diagonalization of *a* formula is not true, without yet saying *which* formula. Let's now consider the diagonalization of $\beta(v_0)$, and let's call it λ:

$$\exists v_0(v_0 \doteq \ulcorner\beta\urcorner \wedge \beta(v_0)) \tag{λ}$$

[6] The concatenation function $\star$ is such that, if s and t are the codes of two expressions, then $s \star t$ is the code of the first expression followed by the second. The numeral function num maps each n to the code of the numeral $\mathbf{n}$. The function $diag$ could have been defined more precisely by first showing that also the logical operations of conjunction and existential quantification are recursive. For more information see Boolos, Burgess, and Jeffrey, (2007, p. 221, §15).

In other symbols, λ is

$$\exists v_0 \big(v_0 \doteq \ulcorner \beta \urcorner \wedge \exists v_1 (\mathbf{Diag}(v_0, v_1) \wedge \neg T(v_1)) \big)$$

This is logically equivalent to $\beta(\ulcorner \beta \urcorner)$, i.e. the result of substituting $\ulcorner \beta \urcorner$ for v_0 in $\beta(v_0)$:

$$\exists v_1 (\mathbf{Diag}(\ulcorner \beta \urcorner, v_1) \wedge \neg T(v_1)) \tag{2}$$

Reading (2) in English, we get something like: "there is a number that has two properties: first, it is the code of the diagonalization of $\beta(v_0)$; second, it is not element of the extension of T". Or, more intuitively: "the diagonalization of β is not true". Interesting enough, the diagonalization of β is precisely λ. Accordingly, λ is logically equivalent to a sentence that says that λ is not true.

We have thus far constructed, within the formal language $\mathscr{L}^t_{pa}$, a sentence saying of itself that it is not true[7]. The next step consists in proving, within **PA**, something about this sentence. Since λ is logically equivalent to (2), we have:

$$\mathbf{PA} \vdash \lambda \leftrightarrow \exists v_1 (\mathbf{Diag}(\ulcorner \beta \urcorner, v_1) \wedge \neg T(v_1)) \tag{3}$$

We do not know, whether λ is a theorem of **PA**. We do know, however, that it is the diagonalization of β, and hence $diag\big(gn(\beta)\big) = gn(\lambda)$. From this, by (1), follows

$$\mathbf{PA} \vdash \forall v_1 (\mathbf{Diag}(\ulcorner \beta \urcorner, v_1) \leftrightarrow v_1 \doteq \ulcorner \lambda \urcorner) \tag{4}$$

That is, $\ulcorner \lambda \urcorner$ is the only closed term satisfying the open formula $\mathbf{Diag}(\ulcorner \beta \urcorner, v_1)$[8]. Simple logic then gives, from (3) and (4):

$$\mathbf{PA} \vdash \lambda \leftrightarrow \exists v_1 (v_1 \doteq \ulcorner \lambda \urcorner \wedge \neg T(v_1)) \tag{5}$$

Since $\exists v_1 (v_1 \doteq \ulcorner \lambda \urcorner \wedge \neg T(v_1))$ is equivalent to $\neg T(\ulcorner \lambda \urcorner)$, we have:

$$\mathbf{PA} \vdash \lambda \leftrightarrow \neg T \ulcorner \lambda \urcorner \qquad \square$$

This is the formal counterpart of the paradoxical Liar sentence: a sentence that is provably equivalent to a sentence saying that its code is not element of the extension of the truth predicate. "But note that $[\lambda]$ is produced by a simple diagonalization construction [...]; and the construction yields a theorem, not a paradox" (Smith,

[7]Whether this sentence "says of itself that it is not true" is not as obvious as one might think. For an insightful discussion about self-reference in arithmetic, see Halbach and Visser, (2014a,b).

[8]Note that (4) is equivalent to the conjunction of $\mathbf{PA} \vdash \mathbf{Diag}(\ulcorner \beta \urcorner, \ulcorner \lambda \urcorner)$ and $\mathbf{PA} \vdash \forall v_1 (\neg (v_1 \doteq \ulcorner \lambda \urcorner) \rightarrow \neg \mathbf{Diag}(\ulcorner \beta \urcorner, v_1))$.

2013, p. 198). The "formal Liar paradox" arises if we want our theory of truth to prove the T-Schema $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ for all sentences $\varphi \in \mathscr{L}^t_{pa}$.

Yet, this is by no means necessary. Kripke (1975) proposes to give up the beloved T-Schema, constructing a partial model for the truth predicate, where both $\lambda \leftrightarrow T\ulcorner\lambda\urcorner$ and $\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$ are neither true nor untrue, i.e. they are undefined. As indicated in the INTRODUCTION, I assume the reader being familiar with Kripke's *Outline*. I omit completely the presentation of his work. Here I shall just state two important features of MFP, described by Kripke (1975, p. 708) as "probably the most natural model for the intuitive concept of truth".

Fact 1.2. MFP verifies the metalinguistic T-Schema, i.e.: for all sentences $\varphi \in \mathscr{L}^t_{pa}$,

$$\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \varphi \iff \langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} T\ulcorner\varphi\urcorner$$
$$\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \neg\varphi \iff \langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \neg T\ulcorner\varphi\urcorner$$

Fact 1.3. In MFP both the Liar sentence λ and the Truth-teller τ are undefined.

2 Towards a New Model

In this section I shall put forward the new response to the Liar antinomy. The gist of my proposal is that 'paradoxical' ought to be treated as a truth value. Liar sentences, according to the present suggestion, do not simply *lack* a truth value. They do possess one: they are *paradoxical*. As has been noted in the INTRODUCTION, the trigger of my considerations will be the difference between the Liar and the Truth-teller. The main goal is to construct a model within which (i) the difference between paradoxical and unparadoxical statements is detected, and (ii) every $\mathscr{L}^t_{pa}$-sentence φ has the same truth value as $T\ulcorner\varphi\urcorner$ (that's the metalinguistic T-Schema).

The plan is as follows: the next subsection contains philosophical arguments: I try to explain why Kripke's proposal is not sufficiently satisfactory as response to the Liar, and why, more generally, his MFP does not adequately model the truth predicate. In addition, I shall explain why 'paradoxical' should be treated as a truth value. The remaining subsections carry out this idea formally.

2.1 Why?

Without aiming to be censorious toward Kripke's proposal, but rather with the intention of further developing his elegant ideas, I think that his construction suffers from two inadequacies, which can (I hope) be removed. A first, minor problem his proposal is confronted with is that using the value 'undefined' for paradoxical sentences does not seem entirely adequate[9] – at least if we adhere to the original meaning attributed to it by Kleene, (1971). A second, major problem is that Kripke's MFP does not model the truth predicate in a satisfactory way. Let me elaborate these reasons in turn.

In both Kleene's logics (the *Strong* and the *Weak*)[10], the value 'undefined' ($\mathfrak{u}$) is not treated on a pair with 'true' (1) and 'false' (0): $\mathfrak{u}$ is not a third *truth value*[11]; it only represents formally the *lack* of truth values. Secondly, and more important for the present purposes, $\mathfrak{u}$ is open to "*arbitrariness* for a classical value": undefined sentences can turn out to be true or false, or can arbitrarily be declared true or false.

Less tersely: as is well known, Kleene introduced the new logics in the study of partial recursive functions, speaking of which he writes (Kleene, 1971, p. 334):

> if when $Q(x)$ is $\mathfrak{u}$, $Q(x) \vee R(x)$ receives the value 1, the decision must (in the general case) have been made in ignorance about $Q(x)$, and in the face of the possibility that, at some stage in the pursuit of the algorithm for $Q(x)$ later than the last one examined, $Q(x)$ might be found to be 1 or to be 0.

He goes on (*ibid.*, p. 335) to observe that 1, 0, and $\mathfrak{u}$ "must be susceptible of another meaning besides (i) 'true', 'false', 'undefined', namely (ii) 'true', 'false', 'unknown (or value immaterial)'. Here 'unknown' is a category, whose value we either do not know or choose for the moment to disregard; and it does not then exclude the other two possibilities 'true' or 'false' "[12].

My question now is: are paradoxical sentences like the Liar open to the same kind of arbitrariness for a classical value? Might these sentences turn out to be true, or false? Can we arbitrarily assign them a truth value? Hardly so. These sentences are paradoxical precisely because the assumption that they are true, or false, generates inconsistencies.

As already remarked, this is a minor problem. One might quite easily change the interpretation of $\mathfrak{u}$ and adjust it as pleased to paradoxes[13]. Nonetheless, the major problem continues to flutter: MFP does not model the truth predicate adequately, as it does not account for the difference between Liar and Truth-teller – this difference having its roots in a peculiarity of T. Let me make this claim precise, by first repeating that the difference between

[9] Some authors have suggested that paradoxes are overdefined (both true and false), and not underdefined (neither true nor false). See, for example, Dunn, (1969, 1976) and Priest, (1979).

[10] See Kleene, (1971, §64).

[11] Kripke (1975, fn 18) stresses the same point.

[12] Other philosophers have also suggested, as reported by van Fraassen, (1966, pp. 482-483), that sentences that are normally taken to be neither true nor false (for instance "the king of France is wise") "are 'don't cares' for ordinary purposes, and there is therefore no reason why we should not arbitrarily assign them some truth value".

[13] For example, Priest, (1979) introduced the so-called 'Logic of Paradox' (*LP*), which has the same truth tables as the Strong Kleene, but the interpretation of the third value is 'true and false', and it is, moreover, a designated value.

$ The sentence marked with a dollar is untrue.

and

€ The sentence marked with a euro is true.

is that one can more or less arbitrarily declare **€** true, or untrue, without stumbling on logical issues; on the contrary, the only way to declare **$** true, or untrue, requires the abandonment of an important principle about truth, i.e. that nothing is both true and untrue. Therefore, doing nothing more and nothing less than describing a simple state of affairs, we can state that

(Fact) *the truth predicate is such that, there are sentences that can* consistently *be in its extension or in its anti-extension; there are sentences that cannot.*

Every theory of truth ought to take **(Fact)** into account[14].

As a matter of fact, in a substantial portion of the *Outline*, Kripke shows how to categorise different kinds of sentence. A sentence is *paradoxical*, e.g., "if it has no truth value in *any* fixed point" (Kripke, 1975, p. 708)[15]. A sentence is *ungrounded* and *unparadoxical*, if it has a truth value in *some* fixed point, different from the minimal one – an example being the Truth-teller. He even emphasises that "the assignment of a truth value to [the Truth-teller] is *arbitrary*" (*ibid.*, p. 709)[16].

The reader might therefore ask, what the point of my objection is – Kripke *does* offer a way to distinguish between paradoxical and simply undefined sentences; Kripke *does* account for the difference between Liars and Truth-tellers. He surely does. But the point is that only within the *metatheory* one can implement that distinction. Only within an informal "metamodel" of the various fixed-point models are we able to differentiate between paradoxical and unparadoxical sentences. The minimal fixed point, which (*repetita iuvant*) is described by Kripke as "probably the most natural model for the intuitive concept of truth" (*ibid.*, p. 708), doesn't see the difference: in this model the Liar and the Truth-teller are both simply undefined.

If I am right, and if the difference between **$** and **€** is due to the peculiarity of T expressed by **(Fact)**[17], then I believe it is justified to maintain the Kripke's model

[14] A similar point is made by Gupta and Belnap, (1993, p. 100): "The essential thing about the Liar appears to be its instability under semantic evaluation: No matter what we hypothesize its value to be, semantic evaluation refutes our hypothesis. A theory of truth ought to capture *this* intuition. It should provide a way of distinguishing sentences that exhibit this behaviour from those that do not, and it should explain *why* certain sentences behave this way".

[15] Kripke considers only *consistent* fixed point, i.e. fixed point where $E \cap A = \emptyset$. So do I.

[16] Halbach, (2014, p. 196) observes that "Kripke's main contribution was not so much the construction of the smallest fixed point [...] but rather his classification of the different consistent fixed points and the discussion of their use for discriminating between ungrounded sentences, paradoxical sentences, and so on".

[17] Are there any other predicates which are akin to T in this respect? One is there for sure: the predicate "is heterological" introduced by Kurt Grelling and Leonard Nelson (see Grelling and Nelson, 1907). In a parallel work, I am trying to extend the solution presented here to handle the Grelling-Nelson paradox too.

is not quite accurate. I believe it is justified to maintain that we should try to find a way to improve it.

Some suggestions have already been made: it is what McGee, (1991, pp. 110-111) calls a 'liberalisation of Kripke's construction', which allows extension and anti-extension of T to overlap. This requires a replacement of a 3-valued logic with a 4-valued logic having both truth value gaps and truth value gluts. The logical-mathematical properties of such a liberalisation have been studied by Woodruff, (1984)[18]. Such systems are of great interest for dialetheists[19]. But for those who do not believe that something can be ever both true and false, they are of little help. I am one of those, and additionally I really do not believe that declaring the Liar both true and false can represent any kind of solution to the paradox. It seems to me that the paradox *is* precisely that some sentence should be both true and false. I can't digress, however, to discuss dialetheism – intriguing though it might be.

2.2 How?

Although I am not an advocate of dialetheism, I subscribe Visser's words, when he says that "[o]ne attractive feature of four valued logic for the study of the Liar Paradox is the possibility of making certain intuitive distinctions [that is: the distinction between Liars and Truth-tellers. L.C.] *within one single model*" (Visser, 1984, pp. 181-182). And that is why I am about to introduce a new 4-valued logic, whose values are: true, false, paradoxical, and undefined. "Why 'paradoxical'?" – the reader might ask. To properly answer this question, I first need to introduce the idea underlying the new interpretation of T.

We all agree (I venture) that an adequate interpretation of the truth predicate ought to have an extension E and an anti-extension A. Now, since (i) I do not want Liar sentences to simply lay outside $E \cup A$ with Truth-teller sentences, and since (ii) I do not want E and A to overlap, I propose to extend Kripke's interpretation of T by adding a third set to it, which will contain those (codes of) sentences that, as stated in **(Fact)**, cannot consistently be contained in E or in A. I shall call this third set (due to lack of imagination) X. In particular: (E, A, X) will be the interpretation of T, the interpretation of $\mathscr{L}_{pa}$ remaining as before, i.e. we let $\mathcal{M}$ be the standard interpretation of $\mathscr{L}_{pa}$. Consequently, $\langle \mathcal{M}, (E, A, X)\rangle$ will be the interpretation of $\mathscr{L}^t_{pa}$ with, informally:

(i) $E = \{gn(\varphi) \mid \varphi \text{ is true}\}$; $A = \{gn(\varphi) \mid \varphi \text{ is untrue}\}$; $X = \{gn(\varphi) \mid \varphi \text{ is paradoxical}\}$;

[18] See also Visser, (1984).

[19] *Dialetheism*, roughly, is the view that there are true contradictions, and a full exposition of it would involve a great deal of technical material that we will not go into here. See Priest and Berto, (2013) for an overview.

(ii) $E \cap A = \emptyset$, $E \cap X = \emptyset$, $A \cap X = \emptyset$;

(iii) $E \cup A \cup X \neq \mathbb{N}$.

And so now the question arises, what truth value sentences like $T\ulcorner\varphi\urcorner$ should have, whenever $gn(\varphi) \in X$. The answer suggested here, unsurprisingly, is that they are paradoxical. Hence, the reason why I am proposing to take 'paradoxical' as a truth value is that I think the best way to formalise **(Fact)** is having a threefold interpretation of T, with extension, anti-extension, and paradox-set. Accordingly, exactly as though we were allowing E and A to overlap, a fourth truth value is needed. And no value but 'paradoxical' seems to properly suit the paradox-set X.

Now, to carry out this project formally, there are above all three things to be done: first, we need a new 4-valued logic to handle the value 'paradoxical'; second, we need rules determining whether a sentence is true, false, paradoxical, or undefined in the partial model $\langle \mathcal{M}, (E, A, X)\rangle$; third, we need a formal definition of (E, A, X).

2.3 The New Logic

2.3.1 Truth Values and their Structure

Let $\mathbb{C}$ be the class of connectives of classical propositional logic. The new 4-valued logic is defined by the structure:

$$\begin{aligned} \mathcal{W} &= \{1, 0, \mathfrak{p}, \mathfrak{u}\} \\ \mathcal{D} &= \{1\} \\ \mathcal{C} &= \{f_c \mid c \in \mathbb{C}\} \end{aligned}$$

where $\mathcal{W}$ is the set of truth values (true, false, paradoxical, undefined), $\mathcal{D}$ the set of the sole designated value, $\mathcal{C}$ the set of truth functions: for every connective $c \in \mathbb{C}$, f_c is the corresponding truth function. That is: if $c \in \mathbb{C}$ is an n-place connective, f_c is a n-place function with inputs and outputs in $\mathcal{W}$.

As usual, one might order the element of $\mathcal{W}$ by the relation $\leq$. Since $\mathfrak{u}$ represents the lack of truth values, we will have: $\mathfrak{u} \leq 1$; $\mathfrak{u} \leq 0$; $\mathfrak{u} \leq \mathfrak{p}$. The decision to be made concerns the new value $\mathfrak{p}$. There are three possibilities. One might argue that 'paradoxical' represents some sense of 'overdefined', in which case we would have $1 \leq \mathfrak{p}$, $0 \leq \mathfrak{p}$. Or one might say that, like $\mathfrak{u}$, $\mathfrak{p}$ stands for another case of 'underdefined', in which case we would have $\mathfrak{p} \leq 1$ and $\mathfrak{p} \leq 0$. Alternatively, one might say, as I shall do here, that it is neither 'overdefined', nor 'underdefined', whence we have: 1, 0, and $\mathfrak{p}$ are not comparable.

This yields a structure $\mathcal{P} = \langle \mathcal{W}, \leq\rangle$, which can be pictured thus:

$\mathcal{P}$

$$\begin{tikzcd} 1 & \mathfrak{p} & 0 \\ & \mathfrak{u} \arrow[ul] \arrow[u] \arrow[ur] & \end{tikzcd}$$

$\mathcal{P}$ is a poset (partially ordered set), since the ordering $\leq$ on $\mathcal{W}$ is a reflexive, transitive, and antisymmetric binary relation.

Definition 2.1 (Consistency and ccpo)**.** Let $P = \langle D, \leq \rangle$ be a poset. Following Visser, (1984, pp. 184-185), define

(a) A subset $A \subseteq D$ is *consistent* iff each $\{x, y\} \subseteq A$ has an upper bound in D.

(b) P is a *complete, coherent partial order* (ccpo), iff every consistent subset $A \subseteq D$ has a supremum.

Proposition 2.2. $\mathcal{P}$ is a ccpo.

Proof. It is easily verified that each consistent pair of elements $\{\mathfrak{u}, 0\}$, $\{\mathfrak{u}, 1\}$, $\{\mathfrak{u}, \mathfrak{p}\} \subseteq \mathcal{W}$ has a supremum in $\mathcal{W}$ (respectively: 0, 1, $\mathfrak{p}$)[20]. $\square$

2.3.2 Truth Tables and Valuation Function

Instead of defining truth functions singularly[21], I shall for simplicity use the truth tables and I shall write the simple connectives $\neg$, $\vee$, $\wedge \ldots$ instead of $f_\neg$, $f_\vee$, $f_\wedge \ldots$ I also write explicitly conjunction, conditional, and biconditional, although they are defined as usual through negation and disjunction.

$\neg$	
1	0
0	1
$\mathfrak{p}$	$\mathfrak{p}$
$\mathfrak{u}$	$\mathfrak{u}$

$\vee$	1	0	$\mathfrak{p}$	$\mathfrak{u}$
1	1	1	1	1
0	1	0	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{p}$	1	$\mathfrak{p}$	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{u}$	1	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$

$\wedge$	1	0	$\mathfrak{p}$	$\mathfrak{u}$
1	1	0	$\mathfrak{p}$	$\mathfrak{u}$
0	0	0	0	0
$\mathfrak{p}$	$\mathfrak{p}$	0	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{u}$	$\mathfrak{u}$	0	$\mathfrak{u}$	$\mathfrak{u}$

[20] Gupta and Belnap, (1993, §2C) study the mathematical properties of complete coherent partial orders, which turn out to be useful in investigating truth in three-valued languages.

[21] For instance:

$$f_\neg(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x = 1 \\ \mathfrak{p} & \text{if } x = \mathfrak{p} \\ \mathfrak{u} & \text{if } x = \mathfrak{u} \end{cases}$$

$\rightarrow$	1	0	$\mathfrak{p}$	$\mathfrak{u}$
1	1	0	$\mathfrak{p}$	$\mathfrak{u}$
0	1	1	1	1
$\mathfrak{p}$	1	$\mathfrak{p}$	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{u}$	1	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$

$\leftrightarrow$	1	0	$\mathfrak{p}$	$\mathfrak{u}$
1	1	0	$\mathfrak{p}$	$\mathfrak{u}$
0	0	1	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{p}$	$\mathfrak{p}$	$\mathfrak{p}$	$\mathfrak{p}$	$\mathfrak{u}$
$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$

Do the tables suit our intuitions about paradoxality? I will discuss this question below, in Discussion 2.4. But before that, let me define the valuation function $\mathcal{V}_{\langle \mathcal{M},(E,A,X)\rangle} : \mathscr{L}^t_{pa} \longrightarrow \{1, 0, \mathfrak{p}, \mathfrak{u}\}$. For the sake of readability, I shall write $\mathcal{V}$ instead of $\mathcal{V}_{\langle \mathcal{M},(E,A,X)\rangle}$.

(a) For atomic $\mathscr{L}_{pa}$-sentences:

$$\mathcal{V}(\varphi) = \begin{cases} 1 & \text{if} \quad \mathcal{M} \models \varphi \\ 0 & \text{if} \quad \mathcal{M} \models \neg\varphi \end{cases}$$

(b) For atomic $\mathscr{L}^t_{pa}$-sentences $T(\mathbf{n})$:

$$\mathcal{V}(T(\mathbf{n})) = \begin{cases} 1 & \text{if} \quad n \in E \\ 0 & \text{if} \quad n \in A \\ \mathfrak{p} & \text{if} \quad n \in X \\ \mathfrak{u} & \text{if} \quad n \notin E \cup A \cup X \end{cases}$$

(c)

$$\mathcal{V}(\neg\varphi) = \begin{cases} 1 & \text{if} \quad \mathcal{V}(\varphi) = 0 \\ 0 & \text{if} \quad \mathcal{V}(\varphi) = 1 \\ \mathfrak{p} & \text{if} \quad \mathcal{V}(\varphi) = \mathfrak{p} \\ \mathfrak{u} & \text{if} \quad \mathcal{V}(\varphi) = \mathfrak{u} \end{cases}$$

(d)

$$\mathcal{V}(\exists v_i \varphi(v_i)) = \begin{cases} 1 & \text{if} \quad \exists n \in \mathbb{N}\,\big(\mathcal{V}\big(\varphi(\mathbf{n})\big) = 1\big) \\ 0 & \text{if} \quad \forall n \in \mathbb{N}\,\big(\mathcal{V}\big(\varphi(\mathbf{n})\big) = 0\big) \\ \mathfrak{p} & \text{if} \quad \text{(see below)} \\ \mathfrak{u} & \text{if} \quad \text{(see below)} \end{cases}$$

The definition for compound sentences containing connectives is given on the basis of the valuation scheme. The definition for quantified sentences is more intricate, so let me explain the process that brought me at the definition presented below.

When does a sentence beginning with a quantifier have semantic value $\mathfrak{p}$? The answer to this question is crucial, since the various paradoxical sentences are exactly quantified sentences. More precisely, they have the form $\exists v_0(\varphi(v_0) \wedge \neg T(v_0))$, where the code of the sentence is the *only* object satisfying the formula $\varphi(v_0)$, so that for all other numbers n, $\varphi(\mathbf{n})$ is false.

Now, the semantic rules determining when a quantified sentence is true or false can be borrowed from the Strong Kleene semantics adopted by Kripke – as I already did in (d). The problem is that a companion definition for paradoxality, namely

$$\mathcal{V}(\exists v_i \varphi(v_i)) = \mathfrak{p} \text{ iff } \exists n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = \mathfrak{p}\right)$$

is evidently inadequate, since for $\varphi(v_0) \equiv T(v_0)$ there is indeed a $n \in \mathbb{N}$ such that $\mathcal{V}(T(\mathbf{n})) = \mathfrak{p}$, but the sentence "something is true" is not paradoxical. Certainly, nonetheless, the condition that there must be a $n \in \mathbb{N}$, such that $\mathcal{V}(\varphi(\mathbf{n})) = \mathfrak{p}$, is a necessary condition – though not sufficient.

A second thought might be

$$\mathcal{V}(\exists v_i \varphi(v_i)) = \mathfrak{p} \text{ iff } \forall n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = \mathfrak{p}\right)$$

This also does not work, since it would not make λ, as presented in Lemma 1.1, paradoxical. Recall that λ is the sentence $\exists v_0(v_0 \doteq \ulcorner\beta\urcorner \wedge \beta(v_0))$. But the formula $v_0 \doteq \ulcorner\beta\urcorner \wedge \beta(v_0)$ is not always paradoxical. Quite the opposite, for each $n \neq gn(\beta)$, it is false. Certainly, nonetheless, the condition that $\mathcal{V}(\varphi(\mathbf{n})) = \mathfrak{p}$ for all $n \in \mathbb{N}$ is a sufficient condition – though not necessary.

Combining now sufficient and necessary conditions, I shall propose the following definition:

(d)

$$\mathcal{V}(\exists v_i \varphi(v_i)) = \begin{cases} 1 & \text{if} \quad \exists n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = 1\right) \\ 0 & \text{if} \quad \forall n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = 0\right) \\ \mathfrak{p} & \text{if} \quad \exists n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = \mathfrak{p}\right) \mathbin{/\!\!\!\wedge} \\ & \qquad \forall m \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{m})\big) = \mathfrak{p} \vee \mathcal{V}\big(\varphi(\mathbf{m})\big) = 0\right) \\ \mathfrak{u} & \text{if} \quad \exists n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = \mathfrak{u}\right) \mathbin{/\!\!\!\wedge} \\ & \qquad \neg\exists n \in \mathbb{N}\left(\mathcal{V}\big(\varphi(\mathbf{n})\big) = 1\right) \end{cases}$$

The universal quantifier is defined as usual thus:

$$\forall v_i(\varphi(v_i)) :\leftrightarrow \neg\exists v_i \neg(\varphi(v_i))$$

To complete the rough description of the new logic, let me add that validity is defined in terms of truth preservation: an inference from Σ to φ is valid iff if for each $\psi \in \Sigma$, $\mathcal{V}(\psi) = 1$ (the sole designated value), then $\mathcal{V}(\varphi) = 1$.

It would be interesting to compare this logic with some of the 4-valued logics already studied in the literature. But this demands a larger discussion than is possible here. I shall only make one quick remark:

Remark 2.3. Disjunctive syllogism (from $\varphi \vee \psi$ and $\neg\varphi$ infer ψ) is not valid in the 4-valued logic called *first degree entailment*. This is due to the fact that the designated values of this logic are 1 and $\mathfrak{b}$ (= 'both'). As an example, assume that $\varphi = \mathfrak{b}$ and $\psi = 0$; then $\neg\varphi, \varphi \vee \psi \not\models_{FDE} \psi$, since both $\neg\varphi$ and $\varphi \vee \psi$ are designated (namely $\mathfrak{b}$), but ψ undesignated. On the contrary, it is easily verified that in the logic just sketched disjunctive syllogism is valid, for the only designated value is 1[22]. ■

Let us now turn on the truth tables. They are (i) *truth-functional*, in the sense that the value of a compound is a function of the values of its immediate components; (ii) *normal*, in the sense that the value of a compound is determined by the classical rules whenever the components have classical value; (iii) *monotonic*, for they preserve the relevant order.

Behind them there are four simple thoughts: first, they are an extension of the Strong Kleene (K_3) – in fact, whenever no component is $\mathfrak{p}$, they are exactly as K_3; second, the value 'paradoxical' behaves exactly like $\mathfrak{u}$ in connection with 1 or 0; third, the connection of $\mathfrak{u}$ and $\mathfrak{p}$ is always undefined; fourth, like K_3, they let classical logic be our guiding light, whenever we have "enough classical information". Classical logic, for instance, tells us that a conjunction is false whenever at least one conjunct is false. Accordingly, if a conjunction has a false conjunct, the whole sentence becomes false, *independently* from the value of the other conjunct.

Discussion 2.4. Do the tables suit our intuitions about paradoxality? Besides the case of negation, it is hard to determine, since we do not utter, in the everyday life, many compound sentences containing paradoxes as components. I shall thus make no claim to the optimality of the chosen scheme. By way of an example, however, consider:

♣ The part before the comma of the sentence marked with a clubs sign is untrue, or 0 = 0 [formalisable as $\lambda \vee \mathbf{0} \doteq \mathbf{0}$].

♠ The part before the comma of the sentence marked with a spade sign is untrue, or 0 = 1 [formalisable as $\lambda \vee \mathbf{0} \doteq \mathbf{1}$].

[22] Whether the disjunctive syllogism is a plus or a minus is controversial. See Priest, (2006, p. 154) for a brief discussion.

Although these sentences are highly artificial, they ought to be taken into account when working with formal languages. The former seems to be true, simply because it is a disjunction containing a true disjunct. And, as already remarked, classical logic ought to be our guiding light, whenever classical information is enough.

The second sentence might give some troubles. According to the tables, it is paradoxical and this choice is prompted by two considerations. The first: the sentence is surely neither true nor undefined. Now, if we assume that ♠ is untrue, then both disjuncts have to be untrue (this implication presupposes, again, to follow classical logic as far as possible). But the part before the comma is untrue if, and only if, it is true. The second consideration: it creates a parallel with K_3 and with the work of Kripke. In fact, within Kripke's framework, ♠ would be undefined, and undefined is the value ascribed to λ. Since in the new framework λ has a new truth value, the whole sentence does get a new value as well. Nonetheless, the idea that the sentence is assigned the value of λ is preserved. ■

We can now move on to the last part of this section.

2.4 The New Interpretation of T

To begin with, I shall exploit Kripke's construction of Mfp: in the new interpretation of T, E and A will be identical to E_∞ and A_∞ (the extension and the anti-extension of T in Mfp). Of interest is the definition of the paradox-set and the differentiation between paradoxical and ungrounded-and-unparadoxical sentences. Before I begin, a quick remark on the choice of letting E and A be identical to E_∞ and A_∞. Whereas Kripke maintains that the minimal fixed point is *probably* the most natural model for the intuitive concept of truth, I go a bit further: Mfp *is* the most natural model for the ordinary truth predicate[23]. In a longer philosophical work I would have defended this claim. But limits in space urges us to move on to the formal definition of X.

Recall the way Kripke defines paradoxical sentences, namely: a sentence is paradoxical if, and only if, it does not have a truth value in any (consistent) fixed point, whereas a sentence is ungrounded and unparadoxical iff it has a truth value in some fixed point, different from the minimal one. Now, one might be tempted to formalise Kripke's characterisation word for word, defining X as the set of all (codes of) sentences that are undefined in every fixed point. Such a definition would make all Liar sentences paradoxical, and all Truth-teller sentences unparadoxical – and these are indeed two desiderata of the new model. But an unpleasant consequence would

[23] My statement ranges over kripkean models.

derive from it. Let τ be a Truth-teller[24]. If I defined X as above, the following, e.g., would hold in the new model:

$$\mathcal{V}\big((\tau \wedge \neg\tau) \vee \lambda\big) = \mathfrak{u} \qquad \mathcal{V}\big(T\ulcorner(\tau \wedge \neg\tau) \vee \lambda\urcorner\big) = \mathfrak{p} \tag{6}$$

$$\mathcal{V}\big((\tau \vee \neg\tau) \wedge \lambda\big) = \mathfrak{u} \qquad \mathcal{V}\big(T\ulcorner(\tau \vee \neg\tau) \wedge \lambda\urcorner\big) = \mathfrak{p} \tag{7}$$

As (6)-left never gets a truth value in any fixed point, it should be element of X, so that (6)-right would be paradoxical in $\langle\mathcal{M}, (E, A, X)\rangle$. Yet, (6)-left is undefined in $\langle\mathcal{M}, (E, A, X)\rangle$, because $\tau \wedge \neg\tau$ is undefined and λ paradoxical[25]. Similarly for (7).

Therefore, I cannot define X this way, for that would mean abandoning the prospect of constructing a model where every sentence φ has the same truth value as $T\ulcorner\varphi\urcorner$. I shall hence posit a different definition.

Kripke, (1975, p. 701) makes the following example: "Suppose we are explaining the word 'true' to someone who does not yet understand it. We may say that we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself". Following this example, I would suggest:

> we are entitled to assert of any sentence that it is paradoxical under the circumstances when we cannot assert the sentence itself, without being led to assert that it is untrue.

This informal picture is obviously meant to characterise truth-related paradoxes, like the Liar or like the example from Kripke, (1975, p. 691), which involves a kind of cross-reference between statements: Jones says

(I) Most of Nixon's assertions about Watergate are false.

Suppose now that Nixon's assertions about Watergate are evenly balanced between the true and the false, except for one problematic case:

(II) Everything Jones says about Watergate is true.

Suppose, in addition, that **(I)** is the only statement of Jones about Watergate. It is easy to verify that we cannot assert **(I)** (or **(II)**), without being led to assert that it is untrue: If we assert **(I)**, we are implying that **(II)** is untrue. But this implies that **(I)** is untrue. Similarly if we deny **(I)**[26].

[24] Whereas Liar sentences have the form $\exists v_0(\varphi(v_0) \wedge \neg T(v_0))$, Truth-teller sentences are $\exists v_0(\varphi(v_0) \wedge T(v_0))$. In both cases, the code of the sentence in the only number satisfying the formula $\varphi(v_0)$.

[25] Notice that, although I haven't yet shown it in details, τ is undefined and λ paradoxical according to the definition of quantified sentences above. See infra, Proposition 3.1, for details.

[26] Paying attention at some details, also Yablo's paradoxical sequence (see Yablo, 1993) could be described in the same manner.

Let me now turn the informal description of paradoxical sentences into a formal definition. I shall define the set X inductively. After that, I shall explain how the formal definition relates to the informal characterisation.

First, let $\zeta(n, S)$ abbreviate

(i) $n = gn(\varphi) \mathbin{/\!\!\!\wedge} \mathbf{PA} \vdash \varphi \leftrightarrow \neg T\ulcorner\varphi\urcorner$; or

(ii) $n = gn(\neg\varphi) \mathbin{/\!\!\!\wedge} gn(\varphi) \in S$; or

(iii) $n = gn(\varphi \vee \psi) \mathbin{/\!\!\!\wedge} \quad (gn(\varphi) \in S \mathbin{\vee\!\!\!\vee} gn(\psi) \in S) \mathbin{/\!\!\!\wedge}$
$\big((gn(\varphi) \in S \Rightarrow gn(\psi) \in S \cup A) \mathbin{/\!\!\!\wedge} (gn(\psi) \in S \Rightarrow gn(\varphi) \in S \cup A)\big)$; or

(iv) $n = gn(\varphi \wedge \psi) \mathbin{/\!\!\!\wedge} \quad (gn(\varphi) \in S \mathbin{\vee\!\!\!\vee} gn(\psi) \in S) \mathbin{/\!\!\!\wedge}$
$\big((gn(\varphi) \in S \Rightarrow gn(\psi) \in S \cup E) \mathbin{/\!\!\!\wedge} (gn(\psi) \in S \Rightarrow gn(\varphi) \in S \cup E)\big)$[27]; or

(v) $n = gn(\exists v_i \varphi(v_i)) \mathbin{/\!\!\!\wedge}$
$\exists\!\!\exists m \in \mathbb{N} \left(gn(\varphi(\mathbf{m})) \in S\right) \mathbin{/\!\!\!\wedge} \forall\!\!\!\forall k \in \mathbb{N} \left(gn(\varphi(\mathbf{k})) \in S \cup A\right)$; or

(vi) $n = gn(T(\mathbf{m})) \mathbin{/\!\!\!\wedge} m \in S$.

This gives rise to an operator Γ on the powerset of natural numbers, which is monotone. It is well known that monotone operators on $\mathcal{P}(\mathbb{N})$ have fixed points. The minimal one will be our set X.

Definition 2.5 (Paradox Operator)**.** The paradox operator $\Gamma : \mathcal{P}(\mathbb{N}) \longrightarrow \mathcal{P}(\mathbb{N})$ is a function on the powerset of $\mathbb{N}$, defined thus:

$$\Gamma(S) = \{n \mid \zeta(n, S)\}$$

Example 2.6. Let $S_0 = \{gn(\mathbf{0} \doteq \mathbf{0})\}$. Then $\Gamma(S_0)$ will first of all contain all n, such that $n = gn(\varphi)$ and $\mathbf{PA} \vdash \varphi \leftrightarrow \neg T\ulcorner\varphi\urcorner$. Moreover, by condition (ii), it will contain all $n = gn(\neg\varphi)$ such that $gn(\varphi) \in S_0$. Now, since the only $gn(\varphi) \in S_0$ is $gn(\mathbf{0} \doteq \mathbf{0})$, $gn(\neg(\mathbf{0} \doteq \mathbf{0}))$ will be the only (code of) sentence obtained through condition (ii); by condition (iii), $\Gamma(S_0)$ will contain all $n = gn(\varphi \vee \psi)$ such that $gn(\varphi) \in S_0$ or $gn(\psi) \in S_0 \ldots$ and so forth. In our case, since $S_0 = \{gn(\mathbf{0} \doteq \mathbf{0})\}$, $\Gamma(S_0)$ will contain sentences like $gn(\mathbf{0} \doteq \mathbf{0} \vee \mathbf{0} \doteq \mathbf{0})$ (because $gn(\mathbf{0} \doteq \mathbf{0}) \in S_0$), or $gn(\mathbf{0} \doteq \mathbf{0} \vee \mathbf{1} \doteq \mathbf{2})$ (because $gn(\mathbf{1} \doteq \mathbf{2}) \in A$) and so on. Obviously, it will not contain sentences like $gn(\mathbf{0} \doteq \mathbf{0} \vee \mathbf{1} \doteq \mathbf{1})$ (because $gn(\mathbf{1} \doteq \mathbf{1}) \notin S_0 \cup A$).

Notice that, since $gn(\lambda) \notin S_0 \cup A \cup E$, sentences like $\lambda \vee \psi$, $\lambda \wedge \psi$ will not be in $\Gamma(S_0)$, regardless of the ψ. However, being λ provably equivalent with $\neg T\ulcorner\lambda\urcorner$, it will

[27] The reader knows that the conjunction symbol is not part of the official language. I include it anyway, to obtain a clearer overview.

be, according to condition (i), in $\Gamma(S_0)$. Consequently, $gn(\lambda \vee \psi)$ or $gn(\lambda \wedge \psi)$ will be in $\Gamma(\Gamma(S_0))$, whenever ψ respects the conditions imposed by the definition. ■

In section 2.2, I have claimed that $X \cap E = \emptyset$ and that $X \cap A = \emptyset$. Clearly, if we start the iteration of Γ as shown in EXAMPLE 2.6, we will not obtain this result. On the other hand, I have also claimed that X is the least fixed point of Γ, which is obtained by starting the sequence with $S_0 = \emptyset$. To show that Γ has a least fixed point, it suffices to show that it is a monotone function on $\mathcal{P}(\mathbb{N})$. After having shown the monotonicity, it will follow from general theory of inductive definitions that Γ has a least fixed point.

Lemma 2.7 (MONOTONICITY). Γ is monotone. That is: for all $S_i, S_j \in \mathcal{P}(\mathbb{N})$,

$$S_i \subseteq S_j \Rightarrow \Gamma(S_i) \subseteq \Gamma(S_j)$$

Proof. Let $S_1 \subseteq S_2$ and assume, towards a contradiction

$$\exists\!\!\exists n \in \mathbb{N}(n \in \Gamma(S_1) \mathrel{/\!\!\!\wedge} n \notin \Gamma(S_2)) \tag{8}$$

Let k be a number obtained through existential elimination. From the assumption that $k \in \Gamma(S_1)$ follows:

(i) $k = gn(\varphi) \mathrel{/\!\!\!\wedge} \mathbf{PA} \vdash \varphi \leftrightarrow \neg T\ulcorner\varphi\urcorner$; or

(ii) $k = gn(\neg\varphi) \mathrel{/\!\!\!\wedge} gn(\varphi) \in S_1$; or

(iii) $k = gn(\varphi \vee \psi) \mathrel{/\!\!\!\wedge} \quad (gn(\varphi) \in S_1 \mathrel{\vee\!\!\!\vee} gn(\psi) \in S_1) \mathrel{/\!\!\!\wedge}$
$\big((gn(\varphi) \in S_1 \Rightarrow gn(\psi) \in S_1 \cup A) \mathrel{/\!\!\!\wedge} (gn(\psi) \in S_1 \Rightarrow gn(\varphi) \in S_1 \cup A)\big)$; or

(iv) $k = gn(\varphi \wedge \psi) \mathrel{/\!\!\!\wedge} \quad (gn(\varphi) \in S_1 \mathrel{\vee\!\!\!\vee} gn(\psi) \in S_1) \mathrel{/\!\!\!\wedge}$
$\big((gn(\varphi) \in S_1 \Rightarrow gn(\psi) \in S_1 \cup E) \mathrel{/\!\!\!\wedge} (gn(\psi) \in S_1 \Rightarrow gn(\varphi) \in S_1 \cup E)\big)$; or

(v) $k = gn(\exists v_i \varphi(v_i)) \mathrel{/\!\!\!\wedge}$
$\exists\!\!\exists n \in \mathbb{N}\big(gn(\varphi(\mathbf{n})) \in S_1\big) \mathrel{/\!\!\!\wedge} \forall\!\!\forall m \in \mathbb{N}\big(gn(\varphi(\mathbf{m})) \in S_1 \cup A\big)$; or

(vi) $k = gn(T(\mathbf{n})) \mathrel{/\!\!\!\wedge} n \in S_1$.

It can be shown that each of (i)-(vi) implies that $k \in \Gamma(S_2)$.

If (i), then trivially $k \in \Gamma(S_2)$.

If (ii), as $S_1 \subseteq S_2$, $gn(\varphi) \in S_2$, and hence $gn(\neg\varphi) \in \Gamma(S_2)$.

For (iii), let me proceed slowly, step by step. First of all, I have to show that:

$$(iii) \quad \Rightarrow \quad (iii)[S_2/S_1] \tag{9}$$

That is: if (iii) is true (viz. if $k = gn(\varphi \vee \psi) \curlywedge (gn(\varphi) \in S_1 \curlyvee gn(\psi) \in S_1) \curlywedge \ldots)$, then also $(iii)[S_2/S_1]$ is verified. Now, since we are assuming (iii), we are assuming in particular that $(gn(\varphi) \in S_1 \curlyvee gn(\psi) \in S_1)$, which implies the second conjunct of $(iii)[S_2/S_1]$, i.e. $(gn(\varphi) \in S_2 \curlyvee gn(\psi) \in S_2)$ (the first conjunct holds anyway). In order to show the third conjunct, I shall conduct a proof by cases: exploiting the assumption that $(gn(\varphi) \in S_1 \curlyvee gn(\psi) \in S_1)$, I shall show that both implies the third conjunct of $(iii)[S_2/S_1]$. In symbols:

$$(gn(\varphi) \in S_1 \curlyvee gn(\psi) \in S_1) \Rightarrow$$
$$\big((gn(\varphi) \in S_2 \Rightarrow gn(\psi) \in S_2 \cup A) \curlywedge (gn(\psi) \in S_2 \Rightarrow gn(\varphi) \in S_2 \cup A)\big) \tag{10}$$

Assume first that $gn(\varphi) \in S_1$. Then $gn(\varphi) \in S_2$, and therefore the second conjunct of (10) is true. To show the first conjunct, notice that from the assumption that $gn(\varphi) \in S_1$ follows that $gn(\psi) \in S_1 \cup A$ and hence that $gn(\psi) \in S_2 \cup A$. This verifies the first conjunct of (10) and concludes the first part of the proof by cases, that is to say: if $gn(\varphi) \in S_1$, then the third conjunct of $(iii)[S_2/S_1]$ is true.

The second part of the proof by cases, which involves the assumption that $gn(\psi) \in S_1$, is exactly the same (*mutatis mutandis*, of course). Hence, if (iii), then $(iii)[S_2/S_1]$ and therefore $k \in \Gamma(S_2)$.

If (iv), then it suffices to substitute E for A in the argument above.

If (v), then there is a $n \in \mathbb{N}$, such that $gn(\varphi(\mathbf{n})) \in S_2$ and for all $m \in \mathbb{N}$, $gn(\varphi(\mathbf{m})) \in S_2 \cup A$. Therefore $gn(\exists v_0 \varphi(v_0)) \in \Gamma(S_2)$.

If (vi), then $n \in S_2$ and hence $gn(T(\mathbf{n})) \in \Gamma(S_2)$.

(8) is therefore false, and the monotonicity of Γ is proved. □

Since Γ is a monotone operator on $\mathcal{P}(\mathbb{N})$, it has a least fixed point.

Lemma 2.8 (Fixed Point)**.** Γ has a minimal fixed point, i.e. there is a set S such that $\Gamma(S) = S$, and for all $S' = \Gamma(S')$, $S \subseteq S'$.

Proof Sketch. Every monotone function $\pi : P \longrightarrow P$ on an inductive poset P[28] has a (unique) least fixed point. Since the paradox operator Γ is a monotone function on the power set of natural numbers, and since $\mathcal{P}(\mathbb{N})$ is an inductive poset, Γ has a least fixed point[29]. □

[28] A poset P is *inductive* (or *chain-complete*) if every chain $S \subseteq P$ has a least upper bound. (Moschovakis, 2006, Def. 6.10, p. 75).

[29] See Moschovakis, (2006, §§6-7), and Moschovakis, (1974, pp. 6-8) for details. The former contains an extensive, yet accessible, analysis of fixed points in general. The latter is a study of inductive definitions.

The informal description of paradoxical sentences stated above is captured by the first clause of the formal definition, viz. $n = gn(\varphi) \mathbin{/\!\!\!\wedge} \mathbf{PA} \vdash \varphi \leftrightarrow \neg T\ulcorner\varphi\urcorner$. It makes sure that the "atomic" paradoxical sentences are elements of X. These sentences, evidently, are not atomic in the usual sense. Nevertheless, they are atomic in the sense that they are the minimum required to yield a paradox. All other clauses are meant to avoid the problem which would have followed from a definition in "Kripke-style"[30]. In other words: their goal is, on the basis of the truth tables presented in § 2.3.2, to assure that in the new model a sentence φ is paradoxical if, and only if, $T\ulcorner\varphi\urcorner$ is paradoxical too. A proof of this claim is contained in the following and last section, which contain the main theorem of the paper.

3 Analysis of the New Model

Let us check whether the model constructed thus far adequately models the truth predicate, and whether it improves the kripkean MFP. First of all, I will show that the Liar gets assigned value $\mathfrak{p}$. Thereafter, I shall prove that the new model verifies the metalinguistic T-Schema.

Proposition 3.1. In $\langle \mathcal{M}, (E, A, X)\rangle$ both λ and $\neg T\ulcorner\lambda\urcorner$ are paradoxical.

Proof. I follow the notation of Lemma 1.1. Since λ is provably equivalent (in **PA**) with $\neg T\ulcorner\lambda\urcorner$, it follows that $gn(\lambda) \in X$ and therefore $\mathcal{V}(T\ulcorner\lambda\urcorner) = \mathfrak{p}$ iff $\mathcal{V}(\neg T\ulcorner\lambda\urcorner) = \mathfrak{p}$.

To prove that $\mathcal{V}(\lambda) = \mathfrak{p}$, as λ is a sentence beginning with a quantifier, namely

$$\exists v_0\big(\underbrace{v_0 \doteq \ulcorner\beta\urcorner \wedge \exists v_1(\mathbf{Diag}(v_0, v_1) \wedge \neg T(v_1))}_{\lambda^-(v_0)}\big)$$

I have to show that there is a $n \in \mathbb{N}$, such that $\lambda^-(\mathbf{n})$ is paradoxical, and that for all $m \in \mathbb{N}$, $\lambda^-(\mathbf{m})$ is either false or paradoxical.

It is clear that for all $m \neq gn(\beta)$, $\lambda^-(\mathbf{m})$ is false. Therefore, I only have to show that $\lambda^-(\ulcorner\beta\urcorner)$ is paradoxical:

$$\begin{gathered}\mathcal{V}(\lambda^-(\ulcorner\beta\urcorner) = \mathfrak{p}) \iff \\ \mathcal{V}\big(\ulcorner\beta\urcorner \doteq \ulcorner\beta\urcorner \wedge \exists v_1(\mathbf{Diag}(\ulcorner\beta\urcorner, v_1) \wedge \neg T(v_1))\big) = \mathfrak{p} \iff \\ \mathcal{V}\big(\exists v_1\underbrace{(\mathbf{Diag}(\ulcorner\beta\urcorner, v_1) \wedge \neg T(v_1))}_{\lambda^{--}(v_1)}\big) = \mathfrak{p}\end{gathered} \tag{11}$$

(11) is easily established. To begin with, for any $m \neq gn(\lambda)$, $\lambda^{--}(\mathbf{m})$ is false, since

[30] A word of warning: I certainly do not mean to suggest that Kripke, in this context, would have defined 'paradoxical' as he did in the *Outline*.

$\mathcal{V}(\mathbf{Diag}(\ulcorner\beta\urcorner, \mathbf{m})) = 0$, for all $m \neq gn(\lambda)$. Furthermore, $\lambda^{--}(\ulcorner\lambda\urcorner)$, i.e.

$$\mathbf{Diag}(\ulcorner\beta\urcorner, \ulcorner\lambda\urcorner) \wedge \neg T\ulcorner\lambda\urcorner \tag{12}$$

is paradoxical, since $\mathcal{V}(\neg T\ulcorner\lambda\urcorner) = \mathfrak{p}$ and $\mathcal{V}(\mathbf{Diag}(\ulcorner\beta\urcorner, \ulcorner\lambda\urcorner)) = 1$. □

I will not show the details for the Truth-teller being undefined, since they are, *mutatis mutandis*, the same.

We can now turn to the main theorem:

Theorem 3.2. (METALINGUISTIC T-SCHEMA) For all $\varphi \in \mathscr{L}^t_{pa}$, the following holds:

$$\mathcal{V}(\varphi) = \mathcal{V}(T\ulcorner\varphi\urcorner)$$

Proof. The proof is quite straightforward, although the details are fairly lengthy. Let me give an outline first: as we know, in MFP every sentence φ has the same truth value as the sentence $T\ulcorner\varphi\urcorner$. LEMMA 3.3 proves that a sentence is true (false) in MFP if, and only if, it has value 1 (0) in $\langle\mathcal{M}, (E, A, X)\rangle$. This gives us the so-called *Nec* (from φ infer $T\ulcorner\varphi\urcorner$) and *Conec* (from $T\ulcorner\varphi\urcorner$ infer φ): a sentence φ has truth value 1 (0) in $\langle\mathcal{M}, (E, A, X)\rangle$ if, and only if, the sentence $T\ulcorner\varphi\urcorner$ has value 1 (0) too. To complete the proof, it remains to be shown that a sentence φ is paradoxical if, and only if, the sentence $T\ulcorner\varphi\urcorner$ is paradoxical as well. This will be done in LEMMA 3.5.

Lemma 3.3. For all $\varphi \in \mathscr{L}^t_{pa}$, the following holds:

$$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\mathrm{sk}} \varphi \Leftrightarrow \mathcal{V}(\varphi) = 1$$
$$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\mathrm{sk}} \neg\varphi \Leftrightarrow \mathcal{V}(\varphi) = 0$$

Proof. The left-to-right direction

$$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\mathrm{sk}} \varphi \Rightarrow \mathcal{V}(\varphi) = 1 \tag{13}$$
$$\langle\mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\mathrm{sk}} \neg\varphi \Rightarrow \mathcal{V}(\varphi) = 0 \tag{14}$$

is evident, since (i) both models have the standard interpretation $\mathcal{M}$ for $\mathscr{L}_{pa}$, (ii) $(E_\infty, A_\infty) = (E, A)$, and (iii) the new logic is exactly like K_3 whenever no conjunct has value $\mathfrak{p}$.

As a shortcut for the right-to-left direction, I will prove that if a sentence has value 1 or 0 in $\langle\mathcal{M}, (E, A, X)\rangle$, then it is not undefined in MFP. It follows that if a sentence has value 1 (0) in $\langle\mathcal{M}, (E, A, X)\rangle$ then it is true (false) in MFP, for it cannot be undefined, nor false (true) – otherwise it would have value 0 (1) in $\langle\mathcal{M}, (E, A, X)\rangle$. Let now

'$\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \varphi$' abbreviate '$\varphi$ is undefined in MFP'. It can be shown that

$$(\mathcal{V}(\varphi) = 1 \vee \mathcal{V}(\varphi) = 0) \;\Rightarrow\; \neg\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \varphi$$

A simple induction verifies the statement.

$\boxed{\varphi \equiv T(\mathbf{n})}$ If $\mathcal{V}(T(\mathbf{n})) = 1$ or $\mathcal{V}(T(\mathbf{n})) = 0$, then $n \in E \cup A$ iff $n \in E_\infty \cup A_\infty$ iff $\neg\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} T(\mathbf{n})$.

$\boxed{\varphi \equiv \neg\psi}$ If $\mathcal{V}(\neg\psi) = 1$ or $\mathcal{V}(\neg\psi) = 0$, then $\mathcal{V}(\psi) = 0$ or $\mathcal{V}(\psi) = 1$. Thus, by i.h., $\neg\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \psi$ iff $\neg\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \neg\psi$.

$\boxed{\varphi \equiv \psi \vee \chi}$ By contraposition, $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \psi \vee \chi$ iff at least one disjunct, say ψ, is undefined and the other, say χ, is not true. By i.h., $\mathcal{V}(\psi) \neq 1$ and $\mathcal{V}(\psi) \neq 0$, and therefore $\mathcal{V}(\psi \vee \chi) \neq 0$. To show that $\mathcal{V}(\psi \vee \chi) \neq 1$, it suffices to show that $\mathcal{V}(\chi) \neq 1$, which follows from the fact that χ is either false or undefined in MFP: if it is false, i.e. if $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \neg\chi$ then $\mathcal{V}(\chi) = 0$, and if $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \chi$, then by i.h. $\mathcal{V}(\chi) \neq 1$. Consequently, $\mathcal{V}(\psi \vee \chi) \neq 1$.

$\boxed{\varphi \equiv \exists v_i(\psi(v_i))}$ By contraposition, $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \exists v_i(\psi(v_i))$ iff there is no $n \in \mathbb{N}$, such that $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \psi(\mathbf{n})$, and for at least some $n \in \mathbb{N}$, $\psi(\mathbf{n})$ is undefined. Hence, by i.h., for some $n \in \mathbb{N}$, $\mathcal{V}(\psi(\mathbf{n})) \neq 0$, and thus $\mathcal{V}(\exists v_i(\psi(v_i))) \neq 0$. To show that $\mathcal{V}(\exists v_i(\psi(v_i))) \neq 1$, assume the contrary to derive a contradiction. $\mathcal{V}(\exists v_i(\psi(v_i))) = 1$ iff $\exists n \in \mathbb{N}(\mathcal{V}(\psi(\mathbf{n})) = 1)$, iff, by i.h., $\neg\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \dashv_{\text{sk}} \psi(\mathbf{n})$. Then either $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \psi(\mathbf{n})$ or $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \neg\psi(\mathbf{n})$. The former implies $\langle \mathcal{M}, (E_\infty, A_\infty)\rangle \models_{\text{sk}} \exists v_i(\psi(v_i))$; the latter implies that $\mathcal{V}(\psi(\mathbf{n})) = 0$, contradicting the assumption. □

LEMMA 3.3 yields the first half of THEOREM 3.2:

Corollary 3.4. (NEC AND CONEC) For all $\varphi \in \mathscr{L}^t_{pa}$, the following holds:

$$\mathcal{V}(\varphi) = 1 \Leftrightarrow \mathcal{V}(T\ulcorner\varphi\urcorner) = 1$$
$$\mathcal{V}(\varphi) = 0 \Leftrightarrow \mathcal{V}(T\ulcorner\varphi\urcorner) = 0$$

Proof. Straightforward consequence of LEMMA 3.3. □

The lemma below completes the proof.

Lemma 3.5. For all $\varphi \in \mathscr{L}^t_{pa}$, the following holds:

$$\mathcal{V}(\varphi) = \mathfrak{p} \;\Leftrightarrow\; \mathcal{V}(T\ulcorner\varphi\urcorner) = \mathfrak{p}$$

Proof. The proof is by induction on the complexity of φ.

$\boxed{\varphi \equiv T(\mathbf{n})}$ $\mathcal{V}(T(\mathbf{n})) = \mathfrak{p}$ iff $n \in X$ iff, by DEFINITION 2.5-(vi), $gn(T(\mathbf{n})) \in X$ iff $\mathcal{V}(T\ulcorner T(\mathbf{n})\urcorner) = \mathfrak{p}$.

Remark 3.6. Notice that we can now use the induction hypothesis $\mathcal{V}(\varphi) = \mathcal{V}(T\ulcorner\varphi\urcorner)$ for all atomic formulas. ■

$\boxed{\varphi \equiv \neg\psi}$ $\mathcal{V}(\neg\psi) = \mathfrak{p}$ iff $\mathcal{V}(\psi) = \mathfrak{p}$ iff, by i.h., $\mathcal{V}(T\ulcorner\psi\urcorner) = \mathfrak{p}$ iff $gn(\psi) \in X$ iff, by DEFINITION 2.5-(ii), $gn(\neg\psi) \in X$ iff $\mathcal{V}(T\ulcorner\neg\psi\urcorner) = \mathfrak{p}$.

Disjunction

$\boxed{\varphi \equiv \psi \vee \chi;\ \Rightarrow}$ $\mathcal{V}(\psi \vee \chi) = \mathfrak{p}$ iff

(A) At least one between ψ and χ, say ψ, is paradoxical.

(B) χ is either false or paradoxical.

From (A),

$$\mathcal{V}(\psi) = \mathfrak{p} \overset{\text{i.h.}}{\Leftrightarrow} \mathcal{V}(T\ulcorner\psi\urcorner) = \mathfrak{p} \Leftrightarrow gn(\psi) \in X \tag{15}$$

Towards a contradiction, assume that $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) \neq \mathfrak{p}$, iff

(i) $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = 1$; or

(ii) $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = 0$; or

(iii) $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = \mathfrak{u}$.

We can rule out (i) and (ii), since, by COROLLARY 3.4, $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = 1(0)$ iff $\mathcal{V}(\psi \vee \chi) = 1(0)$, but we are assuming $\mathcal{V}(\psi \vee \chi) = \mathfrak{p}$. If (iii), then $gn(\psi \vee \chi) \notin X$. On the basis of DEFINITION 2.5-(iii), since we are assuming that $gn(\psi) \in X$, we can argue as follows:

$$(gn(\psi) \in X \curlywedge gn(\psi \vee \chi) \notin X) \Rightarrow gn(\chi) \notin A \cup X \tag{16}$$

It follows that either $gn(\chi) \in E$, or $gn(\chi) \notin E \cup A \cup X$. If the former, then $\mathcal{V}(T\ulcorner\chi\urcorner) = 1 \Leftrightarrow \mathcal{V}(\chi) = 1$, and if the latter, then $\mathcal{V}(T\ulcorner\chi\urcorner) = \mathfrak{u} \overset{\text{i.h.}}{\Leftrightarrow} \mathcal{V}(\chi) = \mathfrak{u}$. Both contradict (B). Hence all (i), (ii), and (iii) deliver a contradiction, from which derives that $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = \mathfrak{p}$.

$\boxed{\varphi \equiv \psi \vee \chi;\ \Leftarrow}$ $\mathcal{V}(T\ulcorner\psi \vee \chi\urcorner) = \mathfrak{p}$ iff $gn(\psi \vee \chi) \in X$, iff

(A) At least one between $gn(\psi)$ and $gn(\chi)$, say $gn(\psi)$, is element of X.

(B) $gn(\chi) \in A \cup X$.

From (A)

$$gn(\psi) \in X \iff \mathcal{V}(T^\ulcorner \psi^\urcorner) = \mathfrak{p} \overset{\text{i.h.}}{\iff} \mathcal{V}(\psi) = \mathfrak{p} \tag{17}$$

Towards a contradiction, assume $\mathcal{V}(\psi \vee \chi) \neq \mathfrak{p}$. Then – again due to COROLLARY 3.4 – $\mathcal{V}(\psi \vee \chi) = \mathfrak{u}$. But if $\mathcal{V}(\psi \vee \chi) = \mathfrak{u}$ and $\mathcal{V}(\psi) = \mathfrak{p}$, then $\mathcal{V}(\chi) = \mathfrak{u}$ and therefore, by i.h., $gn(\chi) \notin A \cup X$, which contradicts (B).

Existential Quantifier

$\boxed{\varphi \equiv \exists v_0 \psi(v_0);\ \Rightarrow}$ $\mathcal{V}(\exists v_0 \psi(v_0)) = \mathfrak{p}$, iff

(A) $\exists n \in \mathbb{N}\,(\mathcal{V}(\psi(\mathbf{n})) = \mathfrak{p})$.

(B) $\forall m \in \mathbb{N}\,(\mathcal{V}(\psi(\mathbf{m})) = \mathfrak{p} \vee \mathcal{V}(\psi(\mathbf{m})) = 0)$.

Using the induction hypothesis, (A) and (B) yield:

(A') $\exists n \in \mathbb{N}\,(gn(\psi(\mathbf{n})) \in X)$.

(B') $\forall m \in \mathbb{N}\,(gn(\psi(\mathbf{m})) \in A \cup X)$.

We derive by DEFINITION 2.5-(v) that $gn(\exists v_0 \psi(v_0)) \in X$, and therefore that $\mathcal{V}(T^\ulcorner \exists v_0 \psi(v_0)^\urcorner) = \mathfrak{p}$.

$\boxed{\varphi \equiv \exists v_i \psi(v_i);\ \Leftarrow}$ $\mathcal{V}(T^\ulcorner \exists v_i \psi(v_i)^\urcorner) = \mathfrak{p}$, iff $gn(\exists v_i \psi(v_i)) \in X$, iff

(A) $\exists n \in \mathbb{N}\,(gn(\psi(\mathbf{n})) \in X)$.

(B) $\forall m \in \mathbb{N}\,(gn(\psi(\mathbf{m})) \in A \cup X)$.

(A) and (B) imply

(A') $\exists n \in \mathbb{N}\,\big(\mathcal{V}(T^\ulcorner \psi(\mathbf{n})^\urcorner) = \mathfrak{p}\big)$.

(B') $\forall m \in \mathbb{N}\,\big(\mathcal{V}(T^\ulcorner \psi(\mathbf{m})^\urcorner) = \mathfrak{p} \vee \mathcal{V}(T^\ulcorner \psi(\mathbf{m})^\urcorner) = 0\big)$.

From (A'), we derive by induction that $\exists n \in \mathbb{N}\,(\mathcal{V}(\psi(\mathbf{n})) = \mathfrak{p})$. From (B'), on the other hand, we derive that $\forall m \in \mathbb{N}\,(\mathcal{V}(\psi(\mathbf{m})) = \mathfrak{p} \vee \mathcal{V}(\psi(\mathbf{m})) = 0)$. Therefore, according to the definition of $\mathcal{V}$, $\mathcal{V}(\exists v_i \psi(v_i)) = \mathfrak{p}$. □

THEOREM 3.2 derives from COROLLARY 3.4 and LEMMA 3.5. □

4 What's next?

There are two questions I didn't address, which lead to an obvious further step. The first is whether the new logic, together with the new model-theoretical framework, may be useful to deal with other paradoxes. Consider, for instance, the Grelling-Nelson paradox (Grelling and Nelson, 1907) involving the predicate "is heterological"[31]. Within the new framework, one might argue that "'heterological' is heterological" is (like the Liar) paradoxical, for 'heterological' cannot consistently be contained in the extension or in the anti-extension of "is heterological", whereas "'autological' is heterological" is (like the Truth-teller) simply undefined.

The second question is how to obtain a proper *theory* of truth, i.e. how an axiomatisation of the new model may look like[32]. Additionally, one might try to add a "Łukasiewicz conditional" to the new logic, to the effect that $f_{\rightarrow}(\mathfrak{p}, \mathfrak{p}) = 1$. Such a conditional could make $\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$ true while both λ and $\neg T\ulcorner\lambda\urcorner$ were still paradoxical. Of course, if one decides to add such a conditional, the interpretation of T must be accordingly modified, in order to preserve the metalinguistic T-Schema. As it is now defined, $gn(\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner) \notin E$, and hence $\mathcal{V}(T\ulcorner\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner\urcorner) \neq 1$. Yet, if in the hypothetical new framework $\mathcal{V}(\lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner) = 1$, then its code better be element of E. This seems to me worthy of study[33]: it does seem right to maintain that the Liar sentence is true if and only if untrue. Would it then not be worthwhile to investigate a theory within which both $T\ulcorner\lambda\urcorner$ and $\neg T\ulcorner\lambda\urcorner$ are paradoxical, but where nonetheless $T\ulcorner\lambda\urcorner \leftrightarrow \neg T\ulcorner\lambda\urcorner$ is true?

[31] Mention should be made at this point of the work of Martin, (1967, 1968), who tries to propose one solution for both Liar and Grelling-Nelson paradoxes.

[32] I guess that an appropriate axiomatisation of the model presented here will result in a system somewhere in the neighbourhood of **PKF** (partial Kripke-Feferman).

[33] A study in a similar direction is due to Field, (2002, 2008), who adds a new conditional to K_3, which is not definable as usual by negation and disjunction.

References

Boolos, George S., John P. Burgess, and Richard C. Jeffrey (2007). *Computability and Logic*. 5th edition. Cambridge (UK): Cambridge University Press.

Burgess, John P. (2011). "Kripke on Truth". In: *Saul Kripke*. Ed. by A. Berger. Cambridge (UK): Cambridge University Press, pp. 141–159.

Cantini, Andrea (1989). "Notes on Formal Theories of Truth". In: *Zeitschrift fur mathematische Logik und Grundlagen der Mathematik* 35, pp. 97–130.

Dunn, J. Michael (1969). "Natural Language versus Formal Language". In: *Association for Symbolic Logic*.

— (1976). "Intuitive Semantics for First-Degree Entailments and 'Coupled Trees'". In: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 29.3, pp. 149–168.

Feferman, Solomon (1991). "Reflecting on incompleteness". In: *Journal of Symbolic Logic* 56.1, pp. 1–49.

Field, Hartry (2002). "Saving the Truth Schema from Paradox". In: *Journal of Philosophical Logic* 31.1, pp. 1–27.

— (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

Fitting, Melvin (1986). "Notes on the mathematical aspects of Kripke's theory of truth". In: *Notre Dame Journal of Formal Logic* 27.1, pp. 75–88.

Grelling, Kurt and Leonard Nelson (1907). "Bemerkungen Zu den Paradoxien von Russell Und Burali-Forti". In: *Abhandlungen Der Fries'schen Schule (Neue Serie)* 2, pp. 300–334.

Gupta, Anil (1982). "Truth and paradox". In: *Journal of Philosophical Logic* 11.1, pp. 1–60.

Gupta, Anil and Nuel Belnap (1993). *The Revision Theory of Truth*. Cambridge (MA): MIT Press.

Halbach, Volker (2014). *Axiomatic Theories of Truth*. Revised edition. Cambridge (UK): Cambridge University Press.

Halbach, Volker and Leon Horsten (2006). "Axiomatizing Kripke's Theory of Truth". In: *The Journal of Symbolic Logic* 71.4, pp. 677–712.

Halbach, Volker and Albert Visser (2014a). "Self-Reference in Arithmetic I". In: *The Review of Symbolic Logic* 7.4, pp. 671–691.

Halbach, Volker and Albert Visser (2014b). “Self-Reference in Arithmetic II”. In: *The Review of Symbolic Logic* 7.4, pp. 692–7121.

Horsten, Leon (2011). *The Tarskian Turn*. Cambridge (MA): MIT Press.

Kleene, Stephen C. (1971). *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kripke, Saul (1975). “Outline of a Theory of Truth”. In: *The Journal of Philosophy* 72.19, pp. 690–716.

Martin, Robert L. (1967). “Toward a Solution to the Liar Paradox”. In: *The Philosophical Review* 76.3, pp. 279–311.

— (1968). “On Grelling’s Paradox”. In: *The Philosophical Review* 77.3, pp. 321–331.

McGee, Vann (1991). *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis (IN): Hackett Publishing.

Moschovakis, Yiannis N. (1974). *Elementary Induction on Abstract Structures*. Mineola (NY): Dover Publications.

— (2006). *Notes on set Theory*. New York (NY): Springer.

Priest, Graham (1979). “The Logic of Paradox”. In: *Journal of Philosophical Logic* 8.1, pp. 219–241.

— (2006). *An Introduction to Non-Classical Logic*. 2nd edition. Cambridge (MA): Cambridge University Press.

Priest, Graham and Francesco Berto (2013). “Dialetheism”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Reinhardt, William N. (1986). “Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth”. In: *Journal of Philosophical Logic* 15.2, pp. 219–251.

Smith, Peter (2013). *An Introduction to Gödel’s Theorems*. 2nd edition. Cambridge (UK): Cambridge University Press.

Van Fraassen, Bas C. (1966). “Singular Terms, Truth-Value Gaps, and Free Logic”. In: *The Journal of Philosophy* 63.17, pp. 481–495.

Visser, Albert (1984). “Four Valued Semantics and the Liar”. In: *Journal of Philosophical Logic* 13.2, pp. 181–212.

Woodruff, Peter W. (1984). “Paradox, Truth and Logic Part I: Paradox and Truth”. In: *Journal of Philosophical Logic* 13.2, pp. 213–232.

Yablo, Stephen (1993). “Paradox Without Self–Reference”. In: *Analysis* 53.4, pp. 251–252.