Fixed-point models for paradoxical predicates

Luca Castaldo

Department of Philosophy
University of Bristol
castaldluca@gmail.com

Abstract

This article discusses paradoxes of properties and classes, such as the Liar paradox, Russell's paradox, Grelling-Nelson paradox. The main goals are (i) to spell out a common principle underlying these paradoxes and (ii) to provide a semantic construction which is general enough to treat them uniformly, and which has explanatory power as to why certain sentences are paradoxical. The models presented below are fixed-point models à la [Kri75]. However, unlike Kripke models, they have enough resources for distinguishing between different kinds of sentences (e.g., liar-like and truth-teller-like sentences). Moreover, unlike other fixed-point models that can account for different paradoxical phenomena, the ones suggested here capture every insight about truth captured by Kripke models, particularly the idea that truth is consistent and grounded in nonsemantic facts.

Keywords. Liar paradox · Russell's paradox · Grelling-Nelson paradox · Truth-theories · Fixed-point semantics · Non-classical logic · Groundedness

1 Introduction

Studying the paradoxes of self-reference, Graham Priest pointed out that "when one meets [them] for the first time, one is struck by the fact that they all appear to be members of a single family, generated by a common underlying principle" [Pri94, p.25]. Whether or not there is a *single* common principle underlying *all* paradoxes of self-reference is not an easy question. However, it seems undeniable that *some* common principle(s) underlie *some* of them. For example, well-known paradoxes of self-reference such as the Liar, the Grelling-Nelson, and Russell's have arguably several common underlying principles, along with the fact that they all involve some kind of self-reference. In particular, each antinomy involves a predicate (truth predicate, heterologicality predicate, and membership relation, respectively) exhibiting a similar paradoxical behaviour. The main goal of this article consists in spelling out a common principle underlying these paradoxes and in providing a semantic construction which is general enough to treat them uniformly. The underlying idea is that we need a theory of paradox, which explains why some statements are paradoxical.

The existing literature on this kind of paradoxes is overwhelming, and it would be impossible to compare one's proposal with all (or even a significant portion of) the others. This article will focus on approaches involving the very powerful and very popular tool known as *fixed-point semantics*, and our starting point will be Kripke's seminal paper *Outline of a theory of truth*.[1] The *Outline* presents an extremely intuitive and extremely natural theory of truth, and it provides a simple solution to the Liar paradox. In a nutshell, using *Strong Kleene Logic* (SK), Kripke describes a general inductive procedure for obtaining a class of *partial* and *consistent* models for a language containing a self-applicable truth predicate Tr.[2] These models are called *fixed-point models* (or just fixed-points). Fixed-point models of [Kri75] are 'consistent', in the sense that they do not satisfy contradictions, and they are 'partial', in the sense that they allow truth value *gaps*, i.e., sentences which are neither true nor false. In particular, paradoxical sentences (such as liar sentences claiming of themselves that they are false) lack a truth value, they are *undefined*. The procedure described by Kripke for obtaining fixed-point models is versatile, and it can be used to provide a solution for a number paradoxes of properties and classes.[3]

Among the worries raised against [Kri75], one concerns a form of expressive weakness his theory is confronted with, and it relates to the way his account distinguishes between paradoxical sentences and sentences which are not paradoxical but are nonetheless pathological. To explain the problem perspicuously, consider the following list of sentences:

1. The number 1 is bigger than the number 0.
2. The number 1 is smaller than the number 0.
3. Sentence (3) is false.
4. Sentence (4) is true.

Intuitively, these are different kinds of sentences. The first sentence is a simple truth; the second sentence is a simple falsity; the third sentence is a paradoxical *liar sentence* (declaring it true or

[1]The literature inspired by [Kri75] is extremely vast. In particular, there is a branch of the literature related to a Kripkean fixed-point semantics that will not be discussed here. It is the branch that investigates the addition of a well-behaved (or arguably so) material conditional $\rightarrow$ to the base language. Extensive research on this has been carried out by Hartry Field. Combining ingredients of a Kripkean semantics with some revision theoretic techniques [GB93], his main goal consisted in defining a "nice" conditional satisfying all instances of the T-Schema, while avoiding revenge phenomena (see in particular [Fie04, Fie08, Fie14]). For a critical evaluation see, e.g., [RW07, Wel11, Wel14], and for similar investigations see e.g. [Bac13, Ros16]. An alternative proposal has been recently suggested and defended by [Lei19], where the added conditional, although not delivering every instance of the T-Schema, improves the behaviour of the Fieldian conditional.

[2]Actually, it would be more appropriate to talk about "Kripke's method for obtaining semantic theories of truth", rather than "Kripke's theory of truth", as Kripke considers different three-valued semantics, without committing for SK. For simplicity, though, in this paper I concentrate on Kripke's theory based on SK.

[3]The seminal papers for the consistent fixed-point semantics in philosophical logic are [Kri75] and [MW75], both concentrating on the Liar paradox and on models for the truth-predicate (the latter applied Zorn's lemma to obtain maximal fixed points). This method was then adapted e.g. by [Mad83] to the case of classes by allowing gaps in the membership relation. As reported by [Fef84], this kind of construction for type-free theories of predication and classes, where the inductive method of building fixed-point models is used, were introduced much earlier by Fitch and Gilmore, e.g. [Fit48], [Gil74], [Fit80]. For an excellent overview on the historical and technical development of the fixed-point semantics, see [Can09, §§4-5].

false yields a contradiction); the fourth sentence is a pathological-but-unparadoxical *truth-teller sentence* (it can consistently and arbitrarily be declared true or false). It is well known that, as it was pointed out by Visser [Vis84], we cannot make these intuitive distinctions in Kripke models. In particular, none of the models of [Kri75], on its own, is sharp enough to distinguish between Liars, Truth-tellers, and simple truths and falsities (cf. [Vis84, pp. 181-182]).[4] In other words, models of [Kri75] do not know that liar sentences are paradoxical.

A solution to this problem can be found in Visser's article: it can be shown that the inductive construction of fixed-point models can be carried out while allowing both truth-value gaps and truth-value *gluts* (i.e., sentences which are both true and false).[5] Specifically, one can retain the inductive structure presented in [Kri75] while using the well-known four-valued *Belnap-Dunn Logic*, also known as *First Degree Entailment* (FDE).[6] FDE-models, being four-valued, are sharp enough to differentiate different kinds of sentences, since they can declare Liars both-true-and-false, Truth-tellers neither-true-nor-false, and simple truths and falsities just true, or just false, respectively. In other words, four-valued models overcome the inadequacy of Kripke's three-valued ones: they know that Liars are paradoxical.

FDE-fixed-point models (and variants thereof) have been explored in depth, from both a mathematical and a conceptual point of view.[7] Much less attention, though, has received the question whether the use of FDE, whilst delivering more expressive models, lets us preserve virtues of and insights captured by Kripkean models. Perhaps, one of the reasons why this question has not been addressed adequately so far is that it might appear to have an obvious answer. For example, in the introductory remarks of "Bilattices and the theory of truth", Fitting remarks that FDE

> loses none of the original insights, since Kleene's strong three-valued logic is a natural sublogic of Belnap's [...], and it makes possible a treatment that has its own intuitive satisfaction [...]; after all, a sentence asserting its own falsehood could be taken to be overdetermined as well as underdetermined.
>
> [Fit89, p. 225]

However, it will be shown that the question whether FDE-models capture every insight captured by SK-models does not have an obvious answer. Indeed, it will be shown that there is a

[4]As it will be explained below, [Kri75]'s differentiation between these different kinds of objects exploits the whole class of fixed-points. However, no *single* model can see the difference between them.

[5]To be clear: the import of Visser's paper goes far beyond solving the expressiveness issue of models of [Kri75]. I am concentrating on this aspect, as it is most relevant for the present purposes. Also, the same solution can be found in Woodruff's [Woo84].

[6]This logic is due to Nuel Belnap and Michael Dunn, appeared in the seminal papers [Bel77],[Dun76].

[7]The work of Visser and Woodruff has been extended further by Leitgeb [Lei99]. Both Visser and Leitgeb work in an abstract algebraic setting, and the latter has shown that one can obtain fixed-point like models for any logic whose algebra reduct is a De Morgan lattice (note that both First Degree Entailment and Strong Kleene logic are De Morgan lattices). Work in this area has also been done, e.g., by Fitting, who has shown that Kripke-style models can be obtained for the algebraic structures known as *bilattices* [Fit89] (again, note that Belnap-Dunn logic is a bilattice). Mention should also be made of related work by Cantini [Can89], where he studies classical models for a language containing two unary predicates Tr and Fa (truth and falsity) whose extensions are allowed to overlap. See also [Can96].

strong sense in which Fitting's remark does not stand up under a closer scrutiny: besides giving up on the consistency of truth (which could already be considered to be a major departure from SK-models), any FDE-model which is sharp enough for distinguishing between Liars and Truth-tellers loses a crucial insight captured by the minimal SK-fixed-point, namely that truth is *grounded* in nonsemantic facts. That truth is grounded in nonsemantic facts essentially means that what is true ultimately rests on whether certain sentences not containing the truth predicate are true.[8] As we shall see, in sharp FDE-models there are ungrounded sentences which are (strictly) true, or (strictly) false. This creates an important gap within a fixed-point semantics approach, and it gives rise to the following

Question. *Can we construct fixed-point models which (i) have enough resources for distinguishing between different kinds of sentences and (ii) can retain the insights about truth captured by* SK*-models, particularly the idea that truth is consistent and grounded in nonsemantic facts?*

This article will provide a positive answer: using a four-valued but consistent semantics, it will introduce fixed-point models satisfying both (i) and (ii).

There are two further aspects concerning the semantic construction to be introduced which are important to emphasize. The first is that it not only differentiates paradoxical from unparadoxical sentences, but it also has explanatory power as to why certain sentences are paradoxical. More precisely, it provides an inductive characterization of *paradoxical instance* of a predicate, according to which whether a sentence is paradoxical will depend on a set of *base paradoxical sentences*. The second aspect is that the new semantics will spell out a common principle underlying a number of paradoxes involving paradoxical predicates. In particular, the analysis undertaken below supports the idea that the reason why, e.g., liar sentences are paradoxical instances of the truth predicate is the same as the reason why, e.g., the Russell class is a paradoxical instance of the membership relation. Via a suitable formalisation of this common underlying principle, it will be possible to treat various paradoxes uniformly.

Plan of the article. The article is structured as follows: §2 introduces the kind of paradoxes I will be dealing with, informally explaining how they are related to each other. §3 presents the consistent and the paraconsistent fixed-point semantics, explaining in more details the problems mentioned above. §4 outlines informally the alternative semantics, and §5 gives it formal expression. §6 contains an analysis and a comparison between the minimal fixed-points of the different approaches. §7 briefly discusses the issues of expressive weakness and revenge phenomena. The main theorems to come are then proved in a technical Appendix.

2 The paradoxes, an informal sketch

In order to obtain a more general description of the family of paradoxical predicates considered in this article, I consider three representative examples, beginning with the Liar paradox. It is assumed that the designated objects of the truth predicate Tr, i.e., the objects to which Tr

[8] For more on groundedness, see [Yab82], [Lei05], [Hal14, §17].

applies, are *sentences*. Paradoxical instances of Tr are liar sentences, that it, sentences of the form

(L) L is not-true.

These sentences cannot be consistently declared true or not-true. Yet why, it may be asked, Liars *cannot* be consistently declared true or not-true? Suppose—one can continue—we do declare L true. *If* we additionally assumed that a sentence is true iff what the sentence says is the case, that is *if* we assumed that

T-Schema For any sentence φ, $\mathrm{Tr}\ulcorner\varphi\urcorner$ and φ are equivalent,[9]

then we would have to say that L is also not-true, because that is what L says. But if we do *not* assume **T-Schema**, then we *could* consistently declare L true, or not-true. More precisely, then, we have that:

Fact 1 Liars, assuming **T-Schema**, cannot be consistently declared true, or not-true.

Isomorphic considerations hold for what will be called here the *membership paradox*.[10] It is assumed that designated objects of the binary membership relation $\in$ are pairs of objects $\langle a, b\rangle$, where b is a collection of objects. I also assume that collections are themselves objects. Now let $\{x \mid \varphi\}$ denote the collection of objects that satisfy φ, and assume that the relation $\in$ is such that,

M-Schema For any property φ, there exists a collection $\{x \mid \varphi\}$ such that, for any object a, "$a \in \{x \mid \varphi\}$" and $\varphi(a)$ are equivalent.

Now consider the collection $r := \{x \mid x \notin x\}$. Assuming **M-Schema**, the following are equivalent:

$$\{x \mid x \notin x\} \in \{x \mid x \notin x\} \text{ and } \{x \mid x \notin x\} \notin \{x \mid x \notin x\}.$$

It follows that r cannot consistently be declared member or not-member of itself. In other words, the pair $\langle r, r\rangle$ is a paradoxical instance of $\in$. However, if we do *not* assume **M-Schema**, nothing prevents us from saying that $r \in r$ or that $r \notin r$.[11] Hence, more precisely:

Fact 2 The collection $r := \{x \mid x \notin x\}$, assuming **M-Schema**, cannot be consistently declared member of itself, or non-member of itself.

[9] The expression $\ulcorner\varphi\urcorner$ is a name for the sentence φ. On the notion of equivalence more will be said in due course.

[10] I avoid using the label 'Russell's paradox', as the latter has a different flavour, namely the paradox shows that naïve comprehension is false: it is not the case that for any property φ there exists the set $\{x \mid \varphi\}$, otherwise we could construct the set $r := \{x \mid x \notin x\}$ and $r \in r$ iff $r \notin r$. In other words, Russell's paradox (primarily) shows that r is not an object of set theory.

[11] Needless to say, it would be odd to claim that $r \in r$ or that $r \notin r$, just as it would be odd to claim that a liar sentence is true, or that it is not-true. In the first case, it would not be clear what we would mean by 'a is a (non-)member of the collection b', and in the latter it would not be clear what we would mean by 'φ is true'.

A slightly different analysis underlies the Grelling-Nelson paradox [GN07]: the reason why we cannot consistently declare the heterologicality predicate heterological or not-heterological is that, by *definition*, a predicate $\varphi(x)$ is heterological iff $\varphi(x)$ is not φ.[12] Hence:

Fact 3 'Heterological', by definition of 'x is heterological', cannot be consistently declared heterological, or not-heterological.

To generalize then, this article deals with those paradoxes involving a predicate P which has *paradoxical instances* among its designated objects. These instances are paradoxical in the sense that they cannot consistently be declared P or not-P. Or more precisely, *assuming* some basic principle(s) about P, there are objects a for which the claims "a is P" and "a is not-P" are equivalent.

3 Fixed-point semantics, and its philosophical issues

This section discusses the two influential solutions to the kind of paradox mentioned above, i.e., the consistent and the paraconsistent fixed-point semantics. We assume the reader being familiar with these solutions.

3.1 Language and Notation

First-order *Peano arithmetic*, PA, will be used background theory of syntax.[13] The language $\mathcal{L}_P$ denotes the language of PA, in the signature $\{\overline{0}, S, +, \times\}$, extended by a k-ary predicate P. The language $\mathcal{L}_{\mathrm{PA}} := \mathcal{L}_P \backslash \{P\}$ is the P-free fragment of $\mathcal{L}_P$. Terms and formulae are generated by closing off under $\neg, \vee, \exists$ ($\wedge, \forall, \rightarrow, \leftrightarrow$ are defined according to classical logic). An $\mathcal{L}_P$-expression is a term or a formula of $\mathcal{L}_{\mathrm{Tr}}$. The expression $\overline{n}$ is the numeral corresponding to the number $n \in \omega$. Fixing a canonical *Gödel numbering* of $\mathcal{L}_P$-expressions, if e is an $\mathcal{L}_P$-expression, the Gödel number (= gn) of e is denoted by $\#e$ and $\ulcorner e \urcorner$ is the term representing $\#e$ in $\mathcal{L}_{\mathrm{PA}}$.

Unless otherwise specified: 'pathological' means 'pathological-but-unparadoxical'; λ and τ are variables ranging over Liars and a Truth-tellers, that is, fixed points of the formulae $\neg\mathrm{Tr}(x)$ and $\mathrm{Tr}(x)$, for Tr a monadic truth-predicate.[14] The standard interpretation for $\mathcal{L}_{\mathrm{PA}}$ is denoted by $\mathbb{N}$ and the set of natural numbers is denoted by ω. Finally, SK and FDE denote, respectively, *Strong Kleene* logic and *First Degree Entailment* (the latter is also know as *Belnap-Dunn* logic).

[12] For instance, 'monosyllabic' is heterological as it is polysyllabic while 'polysyllabic' is autological, as it is polysyllabic. In a formal setting, we understand properties as formulae with one free variable x. To be more precise (but in this article that is unnecessary), properties are expressed by a formula $\varphi(x, \vec{p})$ with a distinguished free variable x, and $\vec{p}$ a (possibly empty) set of other free variables ("parameters").

[13] This is only for simplicity. No result, nor any philosophical consideration, depends on the use of PA and as base theory.

[14] The syntactic shape of Liars and Truth-tellers depends on the chosen framework. For example, if working with an arithmetic base theory such as Robinson's or Peano arithmetic, they can be obtained via weak diagonalization: Given a formula φ, its weak diagonalization is $\exists x(x = \ulcorner \varphi \urcorner \wedge \varphi)$. To obtain a fixed point of $\neg\mathrm{Tr}(x)$ it suffices to weakly diagonalize $\exists y(\mathrm{Diag}(x, y) \wedge \neg\mathrm{Tr}(y))$, where Diag represents, in PA, the primitive recursive function *diag* mapping the code of each formula to the code of its diagonalization.

3.2 Consistent fixed-points

Let $\langle \mathbb{N}, (E, A)\rangle$ be a *partial* structure expanding $\mathbb{N}$ by an interpretation (E, A) for Tr. Call E (A) the (anti-)extension of Tr. Let $\mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle} : \mathcal{L}_{\mathrm{Tr}} \longrightarrow \{1, 0, \mathfrak{u}\}$ be the SK valuation function assigning a value to $\mathcal{L}_{\mathrm{Tr}}$-sentences in $\langle \mathbb{N}, (E, A)\rangle$. Let $\mathrm{St} \subseteq \omega$ be the set of codes of $\mathcal{L}_{\mathrm{Tr}}$-sentences, and set $\mathrm{NSt} := \omega - \mathrm{St}$. Then

Definition 3.1 (SK-Kripke Jump)**.** The SK-*Kripke Jump* is a function $\Phi^{\mathsf{SK}} : \wp(\omega)^2 \longrightarrow \wp(\omega)^2$ on disjoint pairs (E, A) of subsets of ω defined by:

$$\begin{aligned}
\Phi^{\mathsf{SK}}_E(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) = 1\},\\
\Phi^{\mathsf{SK}}_A(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) = 0\} \cup \mathrm{NSt},\\
\Phi^{\mathsf{SK}}(E, A) &:= (\Phi^{\mathsf{SK}}_E(E, A), \Phi^{\mathsf{SK}}_A(E, A)).
\end{aligned}$$

The operator Φ^{SK} can be shown to have fixed-points, i.e., pairs (E, A) such that $(E, A) = \Phi^{\mathsf{SK}}(E, A)$.[15] In particular, there exists a *least* (or minimal)[16] fixed-point (E_μ, A_μ) which is *contained* in every other fixed-point (E_κ, A_κ), in the sense that $E_\mu \subseteq E_\kappa$ and $A_\mu \subseteq A_\kappa$. Structures $\langle \mathbb{N}, (E, A)\rangle$ such that $(E, A) = \Phi^{\mathsf{SK}}(E, A)$ will be called *consistent* (or SK) fixed-point models (or just fixed-points) for $\mathcal{L}_{\mathrm{Tr}}$, as no sentence in this structures is both true and false. That is to say, fixed-points of Φ^{SK} satisfy the condition

$$E \cap A = \varnothing.$$

Using the class of fixed-points of Φ^{SK}, one can differentiate between paradoxical and pathological instances of Tr as follows. Paradoxical instances are undefined *everywhere*, that is, they are outside the interpretation of Tr in every fixed-point (cf. [Kri75, p. 708]). Pathological-but-unparadoxical instances, on the contrary, are true or false *somewhere*, that is, they are inside the interpretation of Tr in some fixed-point, but they are undefined in the minimal fixed-point.[17]

The apparent problem with this characterization is that we need a "metamodel" of the various fixed-point models to differentiate different kinds of sentences. There is no *single* model that can see the difference between paradoxical instances, pathological instances, and simple truths and falsities. For instance, if we take the minimal fixed-point of Φ^{SK}, then it can be observed that this model does not see the difference between Liars and Truth-tellers, as they are all simply undefined and they are all simply outside the interpretation of Tr. And if we take a fixed-point of Φ^{SK} such that (some) Truth-tellers are, say, in the extension of Tr, then this model would not know the difference between such pathological sentences and other sentences which are simply

[15] Indeed, given a *sound* interpretation of Tr, that is an interpretation (E, A) such that both $E \subseteq \Phi^{\mathsf{SK}}_E$ and $A \subseteq \Phi^{\mathsf{SK}}_A$, we can obtain a fixed-point by iterating Φ^{SK} on this pair transfinitely many times (see e.g. [Fit86] for details).

[16] I will often use 'minimal' instead of the more precise, but in the present context less common, 'least' fixed-point.

[17] It is possible to make finer-grained distinctions between sentences. For example, there are sentences which have an *intrinsic* truth-value, or ungrounded sentences which have always the same value in all fixed-points where they have one, but which do not have an intrinsic truth-value. See [Kri75, pp. 708-709] for details.

true.[18]

The minimal fixed-point is usually taken to be particularly interesting. For example, Kripke described the minimal fixed-point of Φ^{SK} as "probably the most natural model for the intuitive concept of truth" [Kri75, p. 708]. Yet, a natural requirement for a natural model of the intuitive concept of truth seems to be that it be able, as we are, to know the difference between paradoxical sentences (culprit of giving rise to the Liar paradox) and pathological ones (more innocuous sentences, not implying any inconsistency).[19] Albert Visser endorsed four valued logics exactly on the basis of such considerations, emphasizing that "[o]ne attractive feature of four valued logic for the study of the Liar Paradox is the possibility of making certain intuitive distinctions *within one single model*" [Vis84, pp. 181-182]. Visser is referring precisely to the distinction between Liars and Truth-tellers (*Samesayers*, in his terminology). So let us have a closer look at how these four-valued models overcome the inadequacy of the three-valued ones.

3.3 Paraconsistent fixed-points

The consistent fixed-point semantics presented in the previous section can be generalized to a paraconsistent one.[20] Paraconsistent (or FDE) fixed-point models for $\mathcal{L}_{\mathrm{Tr}}$ are the same kind of structure as consistent fixed-points, i.e., they are (possibly) partial structures $\langle \mathbb{N}, (E, A)\rangle$ expanding $\mathbb{N}$ with a pair (E, A) interpreting Tr. The difference between SK and FDE fixed-points lies in the interaction between E and A: in FDE models, the intersection between E and A is allowed to be non empty.

More precisely, letting $\mathcal{V}^{\mathsf{FDE}}_{\langle \mathbb{N},(E,A)\rangle} : \mathcal{L}_{\mathrm{Tr}} \longrightarrow \{1, 0, \mathfrak{b}, \mathfrak{u}\}$ be the FDE valuation function assigning a truth-value to $\mathcal{L}_{\mathrm{Tr}}$-sentences in the structure $\langle \mathbb{N}, (E, A)\rangle$, we have

Definition 3.2 (FDE Kripke Jump)**.** The FDE-*Kripke Jump* is a function $\Phi^{\mathsf{FDE}} : \wp(\omega)^2 \longrightarrow \wp(\omega)^2$ on pairs (E, A) of subsets of ω defined by:

$$\begin{aligned}
\Phi^{\mathsf{FDE}}_E(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{FDE}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) \in \{1, \mathfrak{b}\}\},\\
\Phi^{\mathsf{FDE}}_A(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{FDE}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) \in \{0, \mathfrak{b}\}\} \cup \mathrm{NSt},\\
\Phi^{\mathsf{FDE}}(E, A) &:= (\Phi^{\mathsf{FDE}}_E(E, A), \Phi^{\mathsf{FDE}}_A(E, A)).
\end{aligned}$$

[18] Incidentally, let me remark that a similar problem can be ascribed to revision theories of truth [GB93, Her82a, Her82b]. Several models have to be taken into account in order to differentiate Liars, Truth-tellers, and simple truths and falsities. For a comparison between fixed-point semantics and revision theory, see e.g. [Bur86, Kre09, Wel01].

[19] A very similar point is made by [Ros19]. Indeed, Rossi's article appears to have the same motivation as the present one, even though he restricts the analysis to languages containing a truth predicate. His goal is to obtain a single semantic evaluation that can differentiate different kinds of pathological sentences. His framework, however, is very different from the one introduced below. Rossi provides a graph-theoretic analysis of the paradoxes and, based on an isomorphism about and between semantic graphs, he defines a *canonical evaluation* which constitutes the core of his theory. On this evaluation, the value of a sentence is either a numerical value or a *set of equations*. The framework introduce below, on the other hand, is a more familiar Kripkean-style fixed-point semantics, where models are set-theoretic structures built inductively using a monotone operator on the powerset of ω. In any case, considering the similarity of the motivations for this project and that of [Ros19]'s, I will refer to Rossi's paper elsewhere in this article in order to point out some similarities and differences.

[20] For an overview on paraconsistent logics, see e.g. [Pri02, Rip15].

Also the operator Φ^{FDE} can be shown to have fixed-points $(E, A) = \Phi^{\mathsf{FDE}}(E, A)$,[21] and it also has a least fixed-point contained in every other fixed-point.[22] The class of fixed-points of Φ^{FDE}, though, is more diverse than the class of fixed-points of Φ^{SK}. In particular, fixed-points of Φ^{FDE} may but need not be such that $E \cap A \neq \varnothing$, and they may but need not be such that $E \cup A = \omega$.[23] For simplicity, however, I restrict my attention to those models differentiating between paradoxical and pathological sentences, i.e., those having paradoxical instances in $E \cap A$ and pathological sentences outside $E \cup A$. Hence, when I talk about FDE fixed-point models, I mean structures $\langle \mathbb{N}, (E, A)\rangle$ satisfying the conditions

$$E \cup A \neq \omega \quad \text{and} \quad E \cap A \neq \varnothing.$$

In particular, paradoxical sentences such as liar sentences will be in $E \cap A$, whereas pathological ones such as truth-teller sentences will be simply outside $E \cup A$. Consequently, the differentiation between paradoxical and pathological sentences can be provided within single models. FDE fixed-points, in other words, are able to detect paradoxical instances, and to differentiate them from pathological ones.

However, the price we have to pay for the gained sharpness might be too high. One problem this approach has to face has been mentioned in the introduction and it will be discussed in section 6—in the comments following Observation 6.1 and Proposition 6.2. But another evident problem can be mentioned straight away: we would have to accept contradictions. We would have to say that something (at the same time, in the same respect, etc.) is both true and not-true.[24] The dialetheic idea, according to which some contradictions can be true, is not always met with enthusiasm. Saying that something is both true and false, and more generally saying that something is both P and not-P, is a claim that not everyone is willing to accept. Be that as it may, as a decorous discussion of dialetheism would deserve a separate article, it will not be discussed here. It will just be dogmatically assumed that contradictions cannot be true.[25]

4 A new fixed-point semantics, an informal outline

This section outlines informally the new fixed-point semantics, which overcomes the inadequacies of the semantics presented in the previous section. As already mentioned, one goal consists in developing a semantic construction to deal uniformly with paradoxes involving a paradoxical

[21] They can be obtained in the same way as one obtains consistent fixed-points, i.e., it suffices to iterate Φ^{FDE} transfinitely many times on sound interpretations of Tr. Cf. footnote 15.

[22] Actually, the minimal fixed-points of the SK- and of the FDE-Kripke Jump are identical.

[23] For instance, the least fixed-point is such that $E \cup A \neq \omega$ and $E \cap A = \varnothing$.

[24] A caveat is in order: The authors who first introduced and studied paraconsistent fixed-point models, [Woo84] and [Vis84] (see also [Fef84, Can96, Lei99]), have not endorsed dialetheism, as they were primarily interested in the *mathematical* properties of such constructions (in particular: (i) working with FDE produces a complete lattice of fixed-points, whereas using SK only gives a semilattice and (ii) it can be shown that models satisfying gluts are isomorphic to models having gaps). I mention contradictions and dialetheism because of the way I am treating the pair (E, A) interpreting Tr, i.e., E is taken to contain everything which *is* Tr and A everything which *is* not-Tr. Under this reading (which of course is not the only possibility), a model where $E \cap A \neq \varnothing$ is *ipso facto* a model saying that something *is* both Tr and not-Tr.

[25] [PBAG06] contains several articles defending the possibility of true contradictions. See also e.g. [Pri06, Bea09].

predicate P. Specifically, the aim is to obtain models which know the difference between paradoxical and pathological instances of P, and which know why some instances are paradoxical. This section, then, also discusses what qualifies as paradoxical. In particular, I shall spell out a common underlying principle making, e.g., liar sentences paradoxical instances of the truth predicate, the pair $\langle r, r\rangle$ a paradoxical instance of the membership relation, the predicate 'x is heterological' a paradoxical instance of the heterologicality predicate, and so forth.

4.1 A new structure

In classical first-order logic, the interpretation of a predicate P is just a single set, E. E is the extension of P, and it contains everything which is P, anything else being in the complement of E. Accordingly, we have two truth-values, *true* and *false*, and every sentence has exactly one value: $P(t)$—assuming for simplicity that P is a unary predicate—is true (false), precisely if the value of the term t is (not) in E. As we have seen, this simple picture changes with the fixed-point semantics: an interpretation for $\mathcal{L}_P$ is a more complex structure, where the interpretation of P is a pair of sets (E, A). We still have two truth-values,[26] but a different interaction between E and A is then what distinguishes the consistent and the paraconsistent semantics. In the consistent semantics, some sentences may lack a truth-value. In the paraconsistent semantics, not only some sentences may lack a truth-value, but some other may have both truth-values.

The fixed-point structure presented below does not just modify the interaction between extension and anti-extension of P, but it is a new kind of structure altogether: the interpretation of P is not a single set, it is not a pair, but it is a *triple*; a triple of pairwise disjoint sets (E, A, X), where E and A are the usual extension and anti-extension of P, and X is a *paradox-set*. In other words, an interpretation for $\mathcal{L}_P$ is a structure $\langle \mathcal{N}, (E, A, X)\rangle$, where $\mathcal{N}$ interprets $\mathcal{L}_{\mathrm{PA}}$, and (E, A, X) is an interpretation of P. This new kind of interpretation of paradoxical predicates has several virtues, both philosophical and technical. The salient features of these structures will be presented in some details below, after having provided a more formal definition. But let me here first concentrate on the philosophical rationale behind the addition of this third set, and let me mention one important consequence this has on the set of truth-values.

Looking at the set of designated objects of P, we see that it can be partitioned (informally) into four subsets:

(i) the set of objects which are uncontroversially P;

(ii) the set of objects which are uncontroversially not-P;[27]

(iii) the set of objects a which are neither uncontroversially P, nor uncontroversially not-P, and for which the claims "a is P" and "a is not-P" are equivalent;

[26] I am interpreting the values $\mathfrak{u}$ and $\mathfrak{b}$ as representations of underdetermination and overdetermination, respectively. Sentences with value $\mathfrak{u}$ are neither true, nor false, they are underdetermined. Sentences with value $\mathfrak{b}$ are both true and false, they are overdetermined. Cf. [Kri75] and [Bel19].

[27] Arguably, many paradoxical predicates have this property. For example, $0 = 0$ can be taken to be uncontroversially true, my red shirt can be taken to be uncontroversially outside the set of only blue objects, and so on. If there exist predicates P for which it is always controversial to decide whether something is or is not P, then these predicates are left out from the present analysis.

(iv) the set of objects which are neither uncontroversially P, nor uncontroversially not-P, and which can consistently be declared P or not-P.

This taxonomy should not be surprising, as it is implicitly contained in the fixed-point semantics analysed above. However, there is an aspect of this taxonomy I would like to emphasize, namely that there is a sense in which elements of the first three sets (i)-(iii) are on a par with each other, while elements of the fourth set are not.

Call an interpretation $\mathcal{I}$ of P *acceptable* only if it decides correctly uncontroversial cases—that is, it declares P elements of (i) and non-P elements of (ii). Since we are rejecting dialetheism, it can be observed that in *any* acceptable interpretation of P, objects which are uncontroversially (not-)P are in the (anti-)extension of P, and objects which cannot be consistently declared P or not-P are outside $E \cup A$. However, objects which can consistently be declared P or not-P are in E in *some* acceptable interpretations, they are in A in some other, and they are outside $E \cup A$ in some other still. In this sense, one could say that elements of (i)-(iii) are *exactly* related to extension and anti-extension of P, whereas elements of (iv) are *loosely* related to (E, A). And in this sense, elements of (i)-(iii) are on a par with each other, while elements of (iv) are not. This justifies the addition of a third set into the interpretation of P: objects that are related in the same way to P have the same status in the formal interpretation of P, and each category is contained in one set of the triple (E, A, X).[28]

4.2 A new truth-value.

Having extended the fixed-point style interpretation of a predicate in this way, what about the semantic values that sentences can take within these structures? For example, consider a sentence of the from $P(t)$. Clearly, if the value of the term t is an object which is in the extension of P, then we would expect $P(t)$ to be true. And similarly, if the value of t is an object which is in the anti-extension of P, then we would expect $P(t)$ to be false. The questions here are: if the value of t is an object which is in the paradox-set of P, should the sentence $P(t)$ have a truth-value, and if so, what value should it have? The present suggestion is that this kind of sentences do not simply *lack* a truth-value, nor do they posses a combination of the True and the False. These sentences have a different truth-value: they are *paradoxical*.

Paradoxical, in other words, is taken to be a truth-value, and it is not like the value both-true-and-false of the Belnap-Dunn semantics, which was thought of as a formal representation of the presence of inconsistent data.[29] Recall, in fact, that according to Belnap, "[a sentence] is *ontologically* either True or False, and such ontological truth-values will receive their due" [Bel19, p. 46] (emphasis in the original). To the contrary, here the value Paradoxical is taken to be on a par with the values True and False, just as the set X is on a par with E and A. Following Belnap's terminology, the present suggestion is that a sentence is ontologically True,

[28] I should mention that the idea of adding a third set for paradoxical sentences is not new. As an anonymous referee correctly pointed out to me, in the context of semantic theories of truth it was suggested by Rossi [Ros16], although for different purposes and with a different formal development.

[29] Similarly, the value $\mathfrak{u}$ (undefined) only formally represents the lack of truth-values. Compare also [Kri75, footnote 18].

False, or Paradoxical, and such ontological truth-values will receive their due. It should be stressed that the suggestion of treating Paradoxical as a truth-value is a *consequence* of having a triple interpreting a paradoxical predicate, and not the starting point: since elements of E, elements of A, and elements of X are (in a sense) on a par, it seems natural to let the truth-value of sentences involving them to be (in a sense) on a par.

In *Metaphysics* 1011b25, Aristotle famously stated that

> to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true.

Some things are, and some things are not. Investigating paradoxes, we realize that some things cannot consistently be and cannot consistently not-be. So one could say: *To say of what cannot be and cannot not-be that it is, or that it is not, is paradoxical.*

4.3 Paradoxical instances

But how should we define 'paradoxical instance of P'? This is a difficult question, as it seems that intuitions about what qualifies as paradoxical are debatable. Every account of paradoxicality, arguably, will necessarily involve a degree of arbitrariness. Hence no claim about the optimality of what follows will be made. Nonetheless, the inductive characterization of paradoxical instances of P presented below has several virtues, among which that of avoiding some important objections that Anil Gupta raised against Kripke's characterization of paradoxical sentences (see [Gup82, pp. 33-34]).

4.3.1 Base paradoxical instances.

As mentioned, paradoxical instances of P will be defined inductively. We thus need a starting point. In other words: we need a set of *base* paradoxical instances of P, upon which the paradox-set X can then be inductively constructed. The definition of base paradoxical instances will be the formal expression of the informal characterization of paradoxical instances of P given in §2. Recall that these were described as those elements that, assuming a basic principle about P, cannot be consistently declared P, or not-P. This principle is of the form

$$\text{`}a \text{ is } P\text{' and } \pi[a] \text{ are equivalent,}$$

where $\pi[a]$ expresses the condition(s) a has to satisfy in order to be P. Looking more closely at the sentences giving raise to the paradoxes, it becomes evident that the reason why they cannot be consistently declared P or not-P is that they are equivalent to the *negation* of π. Too see this semi-formally, let e.g. Ht be the heterologicality predicate,[30] λ a liar sentence, and $\dot{\epsilon}$ the

[30] It may be worth mentioning that the predicate "x is heterological" is mathematically definable, in Zermelo-Fraenkel set theory as well as in PA (see [Cie02] for details on the definition (and for a remarkable semantic proof of Gödel's second incompleteness theorem via this heterologicality predicate)). The reason why I use a primitive Ht is to make a parallel between paradoxes involving an arbitrary paradoxical predicate P, which may be not arithmetically definable—the truth predicate being an example.

membership relation. By definition,

$$\text{`}\varphi(x)\text{' is heterological iff } \neg\varphi(\ulcorner\varphi(x)\urcorner).$$

This means that here the principle π states that the formula $\varphi(x)$, in order to be Ht, has to satisfy the condition $\neg\varphi(x)$. Yet, we have that

$$\mathrm{Ht}(\ulcorner\mathrm{Ht}(x)\urcorner) \text{ is equivalent to } \mathrm{Ht}(\ulcorner\mathrm{Ht}(x)\urcorner),$$

that is, $\mathrm{Ht}(x)$ satisfies the negation of π. Similarly, whereas according to **T-Schema**

$$\text{for any sentence } \varphi, \mathrm{Tr}\ulcorner\varphi\urcorner \text{ and } \varphi \text{ are equivalent},$$

we have that

$$\mathrm{Tr}\ulcorner\lambda\urcorner \text{ is equivalent to } \neg\lambda.$$

Similarly again, whereas according to **M-Schema**

for any property φ, there exists a collection $\{x \mid \varphi\}$ such that, for any object a, "$a \in \{x \mid \varphi\}$" and $\varphi(a)$ are equivalent.

we have that, for $\varphi \equiv \neg(x \mathrel{\dot{\epsilon}} x)$ and $a := \{x \mid \neg(x \mathrel{\dot{\epsilon}} x)\}$,

$$\underbrace{a \mathrel{\dot{\epsilon}} \{x \mid \varphi\}}_{\{x|\neg(x \dot{\epsilon} x)\} \dot{\epsilon} \{x|\neg(x \dot{\epsilon} x)\}} \qquad \text{is equivalent to} \qquad \underbrace{\neg\varphi(a)}_{\{x|\neg(x \dot{\epsilon} x)\} \dot{\epsilon} \{x|\neg(x \dot{\epsilon} x)\}}$$

In conformity with this pattern, then, base paradoxical instances of P will be defined as those instances a such that 'a is P' and $\neg\pi[a]$ are equivalent. The notion of equivalence between paradoxical instances and the negation of π will be made precise in the next section.

It is important to emphasize that, albeit not encompassing, the informal characterization of section 2 certainly singles out an important class of paradoxical instances of P. Hence, even though this definition of base paradoxical instances will be incomplete, it may be seen as a valid starting point that could then be improved. Additionally, the just mentioned informal characterization provides a general template for arbitrary predicates P, and it can be used for defining fixed-point models for arbitrary languages $\mathcal{L}_P$. Hence, even if for each particular paradoxical predicate one could provide more accurate definitions of base paradoxical instances, the one just suggested here is general enough to be implemented for an arbitrary paradoxical predicate P.[31]

[31] Relatedly, the present construction is compatible with different definitions of the set of base paradoxical instances, it just needs a definition be given. Hence, even though one might find one specific definition of base paradoxical instances of P implausible, this does not speak against the construction as a whole, but just against a particular implementation thereof.

5 Fixed-point models for paradoxical predicates

This section gives formal expression to the ideas presented in the previous. I begin by defining a new semantics, and I then define a class of fixed-point models for the truth-predicate Tr and for the membership relation $\dot{\epsilon}$. After that, it will be shown how to generalize the construction for an arbitrary paradoxical predicate P, thereby emphasizing one important aspect of the construction below: its generality. It will be shown, in fact, that a structurally identical construction can carried out for an arbitrary paradoxical predicate P.

5.1 SP, the *Semantics of Paradox*

Definition 5.1 (SP matrix)**.** The SP matrix $\mathcal{M}^{\mathsf{SP}}$ is a triple $\langle M, D, O\rangle$, where

- $M = \{1, 0, \mathfrak{p}, \mathfrak{u}\}$ is the set of truth values: true, false, paradoxical, undefined;[32]
- $D = \{1\}$ is the set of the single *designated* value;
- $O = \{\dot\neg, \dot\wedge, \dot\vee\}$ is the set of primitive operations defined on M, whose behaviour is represented in Table 1.[33]

Table 1: Primitive operations

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

$\dot\vee$	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	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$

$\dot\wedge$	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	$\mathfrak{u}$	$\mathfrak{u}$

The truth-tables share two properties with those of FDE.[34] First, they are an extension of SK's truth tables, in the sense that whenever no component is $\mathfrak{p}$, they are exactly as SK's truth tables. Second, it can be observed that the interaction between $\mathfrak{p}$ and a classical value is the same as the interaction between $\mathfrak{b}$ and a classical value, in the following sense. Given a binary operation $\dot\circ$, we have that $1 \dot\circ \mathfrak{p} = 1$ iff $1 \dot\circ \mathfrak{b} = 1$ and $1 \dot\circ \mathfrak{p} = \mathfrak{p}$ iff $1 \dot\circ \mathfrak{b} = \mathfrak{b}$. Similarly for the interaction between $\mathfrak{p}$ and $\mathfrak{b}$ with the value 0.[35] The difference between SP and FDE lies in

[32] I am using $\mathfrak{u}$ as a truth value only for technical simplicity. Philosophically, it only represents the lack of truth values.

[33] The operations $\dot\rightarrow$ and $\dot\leftrightarrow$ are defined as usual.

[34] Although they are well known, let us recall that FDE's truth tables are as follows:

$\dot\neg$	
1	0
0	1
$\mathfrak{b}$	$\mathfrak{b}$
$\mathfrak{u}$	$\mathfrak{u}$

$\dot\vee$	1	0	$\mathfrak{b}$	$\mathfrak{u}$
1	1	1	1	1
0	1	0	$\mathfrak{b}$	$\mathfrak{u}$
$\mathfrak{b}$	1	$\mathfrak{b}$	$\mathfrak{b}$	1
$\mathfrak{u}$	1	$\mathfrak{u}$	1	$\mathfrak{u}$

$\dot\wedge$	1	0	$\mathfrak{b}$	$\mathfrak{u}$
1	1	0	$\mathfrak{b}$	$\mathfrak{u}$
0	0	0	0	0
$\mathfrak{b}$	$\mathfrak{b}$	0	$\mathfrak{b}$	0
$\mathfrak{u}$	$\mathfrak{u}$	0	0	$\mathfrak{u}$

[35] Observe that $\mathfrak{u}$ interacts (in the above sense) with classical values just like $\mathfrak{b}$ and $\mathfrak{p}$.

the interaction between nonclassical values. In fact, in FDE the functions $\dot{\wedge}$ and $\dot{\vee}$ on different nonclassical values yield a classical value, namely $\mathfrak{u}\mathfrak{b} = 0$ and $\mathfrak{u} \underset{\cdot}{\vee} \mathfrak{b} = 1$. Not so in SP, where both $\dot{\wedge}$ and $\dot{\vee}$ output $\mathfrak{u}$ if they are given two different nonclassical values as input, that is, $\mathfrak{u}\mathfrak{p} = \mathfrak{u} \underset{\cdot}{\vee} \mathfrak{p} = \mathfrak{u}$. This choice turns out to be convenient (as we shall later see in Proposition 6.2), but it can be justified independently. Following Kleene, one can take $\mathfrak{u}$ to be a value that might change,[36] for example, the undefined component of $\mathfrak{u} \dot{\vee} \mathfrak{p}$ might become 1, 0, or $\mathfrak{p}$, and depending on which value $\mathfrak{u}$ would turn into, the value of $\mathfrak{u} \dot{\vee} \mathfrak{p}$ would be different. This means that we do not yet have enough information to assign a value to the sentence $\mathfrak{u} \dot{\vee} \mathfrak{p}$, and that is why its value is $\mathfrak{u}$. Note that, for instance, $\mathfrak{u} \dot{\vee} 1 = 1$ because independently of what value $\mathfrak{u}$ would turn into, the value of $\mathfrak{u} \dot{\vee} 1$ will always be 1. To put it differently: we assign a truth-value to a sentences if, and only if, we have enough information to do so; otherwise its value remains $\mathfrak{u}$ (which formally represents the lack of truth-values).

Definition 5.2. An SP-*interpretation* of a first-order language $\mathcal{L}_P$ containing a k-ary predicate P is a structure $\langle \mathcal{N}, (E, A, X) \rangle$ satisfying the following conditions

1. $\mathcal{N}$ is an interpretation of $\mathcal{L}_{\text{PA}}$;
2. $E \cup A \cup X \subseteq |\mathcal{N}|^k$;
3. $E \cap A = \varnothing, E \cap X = \varnothing, A \cap X = \varnothing$.[37]

An SP-interpretation is

standard, if $\mathcal{N} = \mathbb{N}$;

partial, if $E \cup A \cup X \subsetneq |\mathcal{N}|^k$;

total, if $E \cup A \cup X = |\mathcal{N}|^k$;

Kripkean, if it is partial and $X = \varnothing$;[38]

classical, if it is total and $X = \varnothing$.

Let: $\mathcal{I}$ abbreviate (E, A, X); $\mathcal{N}$ be a model for $\mathcal{L}_{\text{PA}}$; $\mathcal{L}_P(\mathcal{N})$ be $\mathcal{L}_P$ expanded by distinct constants $\mathrm{a}, \mathrm{b}, \mathrm{c} \ldots$ for elements $a, b, c \ldots$ of $|\mathcal{N}|$;[39] $t^{\mathcal{N}}$ be the value of the closed term $t \in \mathcal{L}_P(\mathcal{N})$ in $\langle \mathcal{N}, \mathcal{I} \rangle$.

[36][Kle71, p. 334-35] writes that "$\mathfrak{t}$, $\mathfrak{f}$ 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'".

[37]The definition of SP-*interpretation* for a language $\mathcal{L}_P$ based on language $\mathcal{L}$ different from $\mathcal{L}_{\text{PA}}$ and/or with more than one predicate P is done in the obvious way, that is: an SP-interpretation is a structure $\langle \mathcal{N}, (E_1, A_1, X_1), \ldots, (E_n, A_n, X_n), \ldots \rangle$ constituted from a total structure $\mathcal{N} := (|\mathcal{N}|, \ldots)$ for $\mathcal{L}$, together with n (possibly partial) predicates $P_i, 1 \leq i \leq n$.

[38]An SP-interpretation is in fact a generalization of the three-valued models defined by [Kri75].

[39]We expand the language to avoid dealing with variable assignments.

Definition 5.3. A *valuation function* $\mathcal{V}^{\mathsf{SP}}_{\langle\mathcal{N},\mathcal{I}\rangle} : \mathrm{St}_{\mathcal{L}_{\mathrm{P}}(\mathcal{N})} \longrightarrow \{1, 0, \mathfrak{p}, \mathfrak{u}\}$, assigning to each sentence $\varphi \in \mathcal{L}_P(\mathcal{N})$ a truth-value in the structure $\langle\mathcal{N},\mathcal{I}\rangle$, is defined as follows (we write $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}$ instead of $\mathcal{V}^{\mathsf{SP}}_{\langle\mathcal{N},\mathcal{I}\rangle}$):

(a)

$$\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(t = s) = \begin{cases} 1 & \text{if } t^{\mathcal{N}} = s^{\mathcal{N}} \\ 0 & \text{if } t^{\mathcal{N}} \neq s^{\mathcal{N}} \end{cases}$$

(b)

$$\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(P(t_1, \ldots, t_k)) = \begin{cases} 1 & \text{if } \langle t_1^{\mathcal{N}}, \ldots, t_k^{\mathcal{N}}\rangle \in E \\ 0 & \text{if } \langle t_1^{\mathcal{N}}, \ldots, t_k^{\mathcal{N}}\rangle \in A \\ \mathfrak{p} & \text{if } \langle t_1^{\mathcal{N}}, \ldots, t_k^{\mathcal{N}}\rangle \in X \\ \mathfrak{u} & \text{if } \langle t_1^{\mathcal{N}}, \ldots, t_k^{\mathcal{N}}\rangle \notin E \cup A \cup X \end{cases}$$

(c)

$$\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\neg\varphi) = \dot{\neg}\, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi)$$

(d) For $\circ = \vee, \wedge$,

$$\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi \circ \psi) = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi)\ \dot{\circ}\ \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\psi)$$

(e)

$$\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\exists v_i \varphi) = \begin{cases} 1 & \text{if } \text{ for some } a \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}\big(\varphi(\mathrm{a})\big) = 1 \\ 0 & \text{if } \text{ for all } a \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}\big(\varphi(\mathrm{a})\big) = 0 \\ \mathfrak{p} & \text{if } \text{ for some } a \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}\big(\varphi(\mathrm{a})\big) = \mathfrak{p} \text{ and} \\ & \quad\ \text{for all } b \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi(\mathrm{b})) \in \{\mathfrak{p}, 0\} \\ \mathfrak{u} & \text{if } \text{ for some } a \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}\big(\varphi(\mathrm{a})\big) = \mathfrak{u} \text{ and} \\ & \quad\ \text{for no } b \in \mathcal{N}, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}\big(\varphi(\mathrm{b})\big) = 1 \end{cases}$$

Remark 5.4.

(i) The interpretation of the quantifier $\exists$ is *disjunctive*, in the following sense. Just as a disjunction $\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_n$ has value 1 iff some disjunct has value 1, the sentence $\exists v\varphi$ has value 1 iff $\varphi(\mathrm{a})$ has value 1 for some element a. Similarly, just as $\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_n$ has value $\mathfrak{p}$ iff some disjunct has value $\mathfrak{p}$ an any other disjunct has value either $\mathfrak{p}$ or 0, the sentence $\exists v\varphi$ has value $\mathfrak{p}$ iff $\varphi(\mathrm{a})$ has value $\mathfrak{p}$ for some element a and value $\mathfrak{p}$ or 0 for any other element b. Similarly for the values 0 and $\mathfrak{u}$.

(ii) As a consequence of Df. 5.3(a) and of the rules governing the connectives represented in TABLE 1, the valuation function behaves classically on the P-free fragment of $\mathcal{L}_P$.

The SP-consequence relation is defined as preservation of designated values:

Definition 5.5. Let Γ, Δ be finite sets of $\mathcal{L}_P$-sentences. Then Δ is a consequence of Γ, symbolically $\Gamma \models_{\mathsf{SP}} \Delta$, iff for all valuation $\mathcal{V}_\mathcal{I}$ holds: if $\mathcal{V}_\mathcal{I}(\gamma) = 1$ for all $\gamma \in \Gamma$, then $\mathcal{V}_\mathcal{I}(\delta) = 1$ for some $\delta \in \Delta$.

Due to connection mentioned between the truth-tables for SK and those for SP, the following observation should come as no surprise:

Observation 5.6. *Let $\langle \mathbb{N}, (E, A, X)\rangle$ be an arbitrary* SP*-interpretation. Then, for all $\varphi \in \mathcal{L}_{\mathrm{Tr}}$*

$$\mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) = 1 \;\; \textit{iff} \;\; \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},(E,A,X)\rangle}(\varphi) = 1$$
$$\mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle}(\varphi) = 0 \;\; \textit{iff} \;\; \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},(E,A,X)\rangle}(\varphi) = 0$$

Proof. Letting $\vec{t} := t_1, \ldots, t_k$, it suffices to notice that $\mathcal{V}^{\mathsf{SK}}_{\langle \mathbb{N},(E,A)\rangle}(P(\vec{t})) = 1$ (0) iff $\langle t_1^{\mathbb{N}}, \ldots, t_k^{\mathbb{N}}\rangle \in E$ (A) iff $\mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},(E,A,X)\rangle}(P(\vec{t})) = 1$ (0), and then continue by a straightforward induction on φ. $\square$

SP-models, then, are closely related to SK-models. They agree on what is true and on what is false. The difference between them is that sentences which are undefined in a SK-model can be paradoxical in a SP-model. This simple observation turns out to be important for a number of reasons. The first is that it implies that the *logic* of SP is simply SK, which is why I am talking about *semantics* of paradox instead of *logic* of paradox. This is the content of the following

Lemma 5.7. $\Gamma \models_{\mathsf{SK}} \Delta$ *iff* $\Gamma \models_{\mathsf{SP}} \Delta$.

Proof. Suppose $\Gamma \models_{\mathsf{SK}} \Delta$. We want to show that if every sentence in Γ has SP-value 1, then some sentence in Δ has SP-value 1. So assume every sentence in Γ has SP-value 1. Then, by Observation 5.6, it follows that every sentence in Γ has SK-value 1, and hence (since we are assuming $\Gamma \models_{\mathsf{SK}} \Delta$) some sentences in Δ has SK-value 1. The conclusion, i.e., that some sentences in Δ has SP-value 1, follows by Observation 5.6. The right-to-left direction is shown in a similar way. $\square$

If by *logic* we understand a notion of consequence relation between sets of sentences, then the logic underlying SP is SK. This means that SP can be seen as a nonstandard definition of Strong Kleene logic, which is more appropriate for dealing with certain semantic paradoxes.

5.2 $\mathcal{L}_{\mathrm{Tr}}$, and truth-related paradoxes.

In this section, we define fixed-point models for the language $\mathcal{L}_{\mathrm{Tr}}$. As mentioned, the interpretation of Tr will be inductively constructed in stages. The extension E and the anti-extension A of Tr will be *grounded* in arithmetical statements, according to the intuition that the status of the claim that a sentence φ is true or false depends on the prior status of φ or $\neg\varphi$.[40] The paradox-set X, following the considerations of §4.3, will depend on base paradoxical instances of Tr. These are defined as those sentences φ such that $\mathrm{Tr}\ulcorner\varphi\urcorner$ and $\neg\varphi$ are equivalent. Or, to put it in a more familiar and equivalent fashion, φ is a base paradoxical instance of Tr iff φ and

[40]Compare [Soa99, p. 181].

$\neg\mathrm{Tr}\ulcorner\varphi\urcorner$ are equivalent. Equivalent *where*, though? What are the scenarios we are interested in? This question contains *de facto* two subquestions, the first of which is whether we take all models for $\mathcal{L}_{\mathrm{PA}}$ into account or only the standard interpretation. Since this does not have any philosophical implication in this context, we restrict our attention to the standard model for the sake of simplicity. The second subquestion is whether we want to consider all possible interpretations of Tr, or only a subclass of those.

One might e.g. argue that in order for a sentence λ to be a Liar-like sentence, it has to be *always* equivalent to $\neg\mathrm{Tr}\ulcorner\lambda\urcorner$, no matter the interpretation of Tr. Another intuition might be that a sentence λ is Liar-like if it is equivalent to $\neg\mathrm{Tr}\ulcorner\lambda\urcorner$, *while assuming* **T-Schema**. Indeed, we have seen that without **T-Schema** in the background, one could not show that declaring λ true, or untrue, yields an inconsistency. Similarly, one might be interested only in those interpretations of Tr such that being untrue and being false (i.e., having a true negation) are coextensive properties,[41] or in those making Tr a compositional predicate, and so forth. Such intuitions seem to be equally plausible, and it goes without saying that there are several others one can think of. In other words, we can restrict the class of relevant interpretations of Tr in various ways, and the stricter the restriction, the larger the class of base paradoxical instances of Tr will get.

One virtue of the framework below is that it is compatible with different characterizations of base paradoxical instances, hence it will be left open which definition is the most appropriate. For simplicity, I present only two possible definitions, i.e., I consider only two classes of interpretations of Tr, leaving other candidates for future research. The major difference between the two options suggested below concerns the status of some paradigmatic sentences, such as: the McGee sentence μ saying that there is an n such that n iterations of Tr on μ yields a not-true sentence; a Curry sentence κ, saying that either κ is not-true or some absurdity (like $0 \neq 0$) holds; Liar cycles $\lambda_1, \lambda_2, \ldots, \lambda_n$ where each Liar λ_i says of the next Liar λ_{i+1} that it is not-true, and the last Liar λ_n says that the first Liar λ_1 is not-true.[42] As it will turn out, Df. 5.10 implies that both μ and Liar-cycles are *paradoxical*, whereas Df. 5.11 implies that they are *undefined*.

Let me begin by introducing a notion of equivalence over $\mathbb{N}$, which holds between two sentences precisely if they have the same value in every expansion of the standard model:

Definition 5.8 (Equivalence over $\mathbb{N}$)**.** Two sentences φ and ψ of $\mathcal{L}_{\mathrm{Tr}}$ are *equivalent over* $\mathbb{N}$, symbolically $\varphi \equiv^{\mathbb{N}}_{\mathsf{eq}} \psi$, iff $\mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{N},\mathcal{I}\rangle}(\varphi) = \mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{N},\mathcal{I}\rangle}(\psi)$, for all SP-expansions $\mathcal{I} = (E, A, X)$ of $\mathbb{N}$.

Let now an interpretation of Tr be *adequate* iff, for all sentences $\varphi \in \mathcal{L}_{\mathrm{Tr}}$, $\#\varphi \in Y$ iff $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in Y$, for $Y = E, A, X$,[43] and let $\mathcal{A} := \{\mathcal{I} \mid \mathcal{I} \text{ is an adequate interpretation of Tr}\}$. Then

[41] Observe that if we do not assume that being untrue and being false are coextensive properties, we cannot argue that a falsehood-teller, that is, a sentence ϕ equivalent to $\mathrm{Tr}\ulcorner\neg\phi\urcorner$ cannot be consistently declared true or untrue. We could only conclude that, assuming **T-Schema**, that both ϕ and $\neg\phi$ are true.

[42] For a formal definition of these sentences, see Appendix, Proposition 5.16.

[43] Note that this is *not* equivalent to saying that the interpretation is such that, for all φ, φ and $\mathrm{Tr}\ulcorner\varphi\urcorner$ have the same value. For instance, the interpretation $\mathcal{I}$ defined by $E := \{\#\mathrm{Tr}^i\ulcorner\overline{0} = \overline{1}\urcorner \mid i \in \omega\}$, $A := \varnothing =: X$ is such that $\#\varphi \in Y$ iff $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in Y$, for $Y = E, A, X$, but $\mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{N},\mathcal{I}\rangle}(\overline{0} = \overline{1}) = 0$ and $\mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{N},\mathcal{I}\rangle}(\mathrm{Tr}\ulcorner\overline{0} = \overline{1}\urcorner) = 1$. An adequate interpretation can be seen as one which "internalizes" the principle expressed by **T-Schema**.

Definition 5.9 (Equivalence over adequate expansions of $\mathbb{N}$)**.** Two sentences φ and ψ of $\mathcal{L}_{\mathrm{Tr}}$ are *equivalent over adequate expansions of* $\mathbb{N}$, symbolically $\varphi \equiv^{\mathcal{A}}_{\mathsf{eq}} \psi$, iff $\mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},\mathcal{I}\rangle}(\varphi) = \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},\mathcal{I}\rangle}(\psi)$, for all $\mathcal{I} \in \mathcal{A}$.

These two notions of equivalence yield two different definitions of base paradoxical instances of Tr, namely:

Definition 5.10 (Base Liars)**.** The class Λ of *base paradoxical instances* of Tr contains all sentences λ, such that:

$$\lambda \equiv^{\mathcal{A}}_{\mathsf{eq}} \neg\mathrm{Tr}\ulcorner\lambda\urcorner.$$

Definition 5.11 (Base Liars, alternative definition)**.** The class $\Lambda^\star$ of *alternative base paradoxical instances* of Tr contains all sentences λ, such that:

$$\lambda \equiv^{\mathbb{N}}_{\mathsf{eq}} \neg\mathrm{Tr}\ulcorner\lambda\urcorner.$$

Base Truth-tellers are defined similarly, i.e.:

Definition 5.12 (Base Truth-tellers)**.** The class Θ of *base pathological instances* of Tr contains all sentences τ, such that

$$\tau \equiv^{\mathcal{A}}_{\mathsf{eq}} \mathrm{Tr}\ulcorner\tau\urcorner.$$

The class $\Theta^\star$ of *alternative base pathological instances* of Tr contains all sentences τ, such that

$$\tau \equiv^{\mathbb{N}}_{\mathsf{eq}} \mathrm{Tr}\ulcorner\tau\urcorner.$$

Note that $\Lambda^\star \subseteq \Lambda$ and $\Theta^\star \subseteq \Theta$. For simplicity, then, in what follows I often refer only to Λ and Θ. Having defined the necessary atomic blocks, we can now move on to defining a monotone SP-Kripke Jump, whose fixed-points will serve as interpretations of Tr.

Definition 5.13 (SP Kripke Jump)**.** The SP-*Kripke Jump* is a function $\Phi^{\mathsf{SP}} : \wp(\omega)^3 \longrightarrow \wp(\omega)^3$ on pairwise disjoint triples $(E, A, X) = \mathcal{I}$ of subsets of ω defined by:

$$\begin{aligned}
\Phi^{\mathsf{SP}}_E(\mathcal{I}) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},\mathcal{I}\rangle}(\varphi) = 1\},\\
\Phi^{\mathsf{SP}}_A(\mathcal{I}) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},\mathcal{I}\rangle}(\varphi) = 0\} \cup \mathrm{NSt},\\
\Phi^{\mathsf{SP}}_X(\mathcal{I}) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{N},\mathcal{I}\rangle}(\varphi) = \mathfrak{p}\} \cup \Lambda,\\
\Phi^{\mathsf{SP}}(\mathcal{I}) &:= (\Phi^{\mathsf{SP}}_E(\mathcal{I}), \Phi^{\mathsf{SP}}_A(\mathcal{I}), \Phi^{\mathsf{SP}}_X(\mathcal{I})).
\end{aligned}$$

Like Φ^{SK} and Φ^{FDE}, the operator Φ^{SP} can be shown to be monotone and to have fixed-points. To express this more precisely, given two SP interpretations $\mathcal{I}$ and $\mathcal{I}'$ of Tr, let $\mathcal{I} \leq \mathcal{I}'$ be defined as $E \subseteq E'$, $A \subseteq A'$, and $X \subseteq X'$. Then

Lemma 5.14 (Monotonicity)**.** Φ^{SP} *is monotone, i.e.,* $\Phi^{\mathsf{SP}}(\mathcal{I}) \leq \Phi^{\mathsf{SP}}(\mathcal{I}')$ *whenever* $\mathcal{I} \leq \mathcal{I}'$.

Proof. By induction on φ show that, whenever $\mathcal{I} \leq \mathcal{I}'$, for all $\varphi \in \mathcal{L}_{\mathrm{Tr}}$ holds that

$$\text{if } \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi) \in \{1, 0, \mathfrak{p}\}, \text{ then } \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\varphi) = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}'}(\varphi).$$

Φ^{SP}'s monotonicity then easily follows. □

As usual, Φ^{SP}'s monotonicity can be exploited to show the existence of fixed-points, and in particular the existence of a least fixed-point contained in every other fixed-point:

Theorem 5.15.

(a) *Φ^{SP} has a least fixed-point $\mathcal{I}_\mu$ which is contained in every other fixed-point. Moreover, $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(\lambda) = \mathfrak{p}$ for all $\lambda \in \Lambda$ and $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(\tau) = \mathfrak{u}$ for all $\tau \in \Theta$.*

(b) *Any fixed-point $\mathcal{I}_\kappa$ of Φ^{SP} is such that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(\lambda) = \mathfrak{p}$ for all $\lambda \in \Lambda$.*

(c) *Any fixed-point $\mathcal{I}_\kappa$ of Φ^{SP} is such that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(\varphi) = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(\mathrm{Tr}\ulcorner\varphi\urcorner)$, for all $\varphi \in \mathcal{L}_{\mathrm{Tr}}$.*

Proof. See Appendix for a detailed proof, where it is shown how to approach fixed-points from below. Here I only observe that, since the Jump is defined only for pairwise disjoint triples, it is important to verify that the set Λ does not intersect the extension or the anti-extension of Tr at any stage in the transfinite sequence leading to a fixed-point. □

We postpone the analysis of the minimal fixed-point of Φ^{SP} to the next section, where it will be compared with the minimal fixed-points of the consistent and paraconsistent semantics. Let me conclude this section by stating more precisely that, as already mentioned, the main difference between Λ and $\Lambda^\star$ concerns the status of some paradigmatic sentences such as the McGee sentence, Curry sentences and Liar cycles. So let $\Phi^{\mathsf{SP}^\star}$ be defined just as Φ^{SP} but with $\Lambda^\star$ instead of Λ, and let $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$ be the minimal fixed-point of $\Phi^{\mathsf{SP}^\star}$. Then

Proposition 5.16. *(i) A Curry sentence κ is paradoxical in both $\mathcal{I}^{\mathsf{SP}}_{\mu}$ and $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$; (ii) the McGee sentence μ is paradoxical in $\mathcal{I}^{\mathsf{SP}}_{\mu}$ and undefined in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$; (iii) liar cycles are paradoxical in $\mathcal{I}^{\mathsf{SP}}_{\mu}$ and undefined in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$.*

Proof. See Appendix. □

5.3 $\mathcal{L}_{\dot\epsilon}$, and the membership paradox.

Let me begin by emphasising that the goal here is not to provide a model for a theory of sets, classes, or collections. The goal only consists in showing how to define SP-interpretations for a binary membership relation $\dot\epsilon$ with the same method used to construct SP-interpretations for Tr. That is, the aim is to define fixed-point models that (i) know that the pair $\langle r, r\rangle$ is a paradoxical instance of $\dot\epsilon$ and (ii) are such that the value of $t \mathbin{\dot\epsilon} \{x \mid \varphi\}$ is the same as $\varphi[t/x]$ for all $\varphi \in \mathcal{L}_{\dot\epsilon}$.[44]

[44] For a similar construction, see e.g. [Bra71], [Fef84], [Res17]. In [Bra71], so called *class terms* are part of the language.

For the purposes of this section, assume we have a single device for both naming and abstraction for $\mathcal{L}_{\dot\epsilon}$.[45] In particular, for any formula $\varphi \in \mathcal{L}_{\dot\epsilon}$ with free variables among $\vec{x}, \vec{y}$, $(\ulcorner\varphi\urcorner, \vec{y})$ serves as an operation in $\mathcal{L}$ which abstracts $\vec{x}$ treating $\vec{y}$ as parameters. Define:

$$\varphi[\hat{x}_1, \ldots, \hat{x}_k, y_1, \ldots, y_n] = (\ulcorner\varphi\urcorner, y_1, \ldots, y_n)$$

In particular, for $k = 1$, I write (possibly $n = 0$)

$$\{x \mid \varphi(x, y_1, \ldots, y_n)\} = \varphi[\hat{x}, y_1, \ldots, y_n]$$

The x_i's are considered bound and may be renamed by other bound variables. Let $\mathbb{V}$ be the standard interpretation of $\mathcal{L}$, and let its domain contains codes for $\{x \mid \varphi\}$ for any formula φ of $\mathcal{L}_{\dot\epsilon}$.[46]

As above, base paradoxical instances of $\dot\epsilon$ are defined as those pairs $\langle t, s\rangle$ such that the claim $t \mathrel{\dot\epsilon} s$ is equivalent to the negation of the principle about $\dot\epsilon$, that is the **M-Schema**. We just consider the definition taking all interpretations of $\dot\epsilon$ into account, omitting parameters for readability:

Definition 5.17 (Paradoxical and pathological instances)**.** The class R of paradoxical instances of $\dot\epsilon$ contains all pairs $\langle t, \{x \mid \varphi\}\rangle$ such that:

$$t \mathrel{\dot\epsilon} \{x \mid \varphi\} \;\equiv^{\mathbb{V}}_{\mathsf{eq}}\; \neg\varphi[t/x].$$

The class S of pathological but unparadoxical instances of $\dot\epsilon$ contains all pairs $\langle t, \{x \mid \varphi\}\rangle$ such that:

$$t \mathrel{\dot\epsilon} \{x \mid \varphi\} \;\equiv^{\mathbb{V}}_{\mathsf{eq}}\; \varphi[t/x].$$

Observe that $\langle r, r\rangle \in$ R. In fact, $r \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\}$ is equivalent to $\neg\neg(x \mathrel{\dot\epsilon} x)[r/x]$, that is to say, $r \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\}$ is equivalent to the formula obtained from the negation of $\neg(x \mathrel{\dot\epsilon} x)$ by replacing the free variable x by r. More perspicuously, note that

$$\mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{V},\mathcal{I}\rangle}(\underbrace{\{x \mid \neg(x \mathrel{\dot\epsilon} x)\} \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\ \}}_{r \,\dot\epsilon\, \{x \mid \neg(x \,\dot\epsilon\, x)\}}) = \mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{V},\mathcal{I}\rangle}(\underbrace{\{x \mid \neg(x \mathrel{\dot\epsilon} x)\} \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\}}_{(x \,\dot\epsilon\, x)[r/x]})$$

and since every sentence φ is equivalent to $\neg\neg\varphi$, we get

$$\mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{V},\mathcal{I}\rangle}(\underbrace{\{x \mid \neg(x \mathrel{\dot\epsilon} x)\} \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\ \}}_{r \,\dot\epsilon\, \{x \mid \neg(x \,\dot\epsilon\, x)\}}) = \mathcal{V}^{\mathsf{SP}}_{\langle\mathbb{V},\mathcal{I}\rangle}(\underbrace{\{x \mid \neg(x \mathrel{\dot\epsilon} x)\} \mathrel{\dot\epsilon} \{x \mid \neg(x \mathrel{\dot\epsilon} x)\}}_{\neg\neg(x \,\dot\epsilon\, x)[r/x]})$$

In other words, the pair $\langle r, r\rangle$ is equivalent to the negation of **M-Schema**, which in turn implies that $\langle r, r\rangle$ is a paradoxical instance of $\dot\epsilon$.

[45][Fef84] defines a precise methodology for such a device. For details, which are not important here, the reader is referred to Feferman's article.

[46]See [Fef84] for details.

Let Fml $\subseteq |\mathbb{V}|$ consist of all (codes of) $\{x \mid \varphi\}$ for any formula $\varphi(x)$ with at most x free,[47] and let NFml be the set of pairs $\langle t^{\mathbb{V}}, l\rangle$ for $l \notin$ Fml.

Definition 5.18 (Epsilon Jump)**.** The *epsilon Jump*, Γ^{SP}, is a function on pairwise disjoint triples $(E, A, X) = \mathcal{I}$ defined by:

$$\begin{aligned}
\Gamma^{\mathsf{SP}}_E(\mathcal{I}) &:= \big\{\langle t^{\mathbb{V}}, \{x \mid \varphi\}\rangle \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{V}, \mathcal{I}\rangle}(\varphi[t/x]) = 1\big\}\\
\Gamma^{\mathsf{SP}}_A(\mathcal{I}) &:= \big\{\langle t^{\mathbb{V}}, \{x \mid \varphi\}\rangle \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{V}, \mathcal{I}\rangle}(\varphi[t/x]) = 0\big\} \cup \mathrm{NFml}\\
\Gamma^{\mathsf{SP}}_X(\mathcal{I}) &:= \big\{\langle t^{\mathbb{V}}, \{x \mid \varphi\}\rangle \mid \mathcal{V}^{\mathsf{SP}}_{\langle \mathbb{V}, \mathcal{I}\rangle}(\varphi[t/x]) = \mathfrak{p}\big\} \cup \mathrm{R}\\
\Gamma^{\mathsf{SP}}(\mathcal{I}) &:= (\Gamma^{\mathsf{SP}}_E(\mathcal{I}), \Gamma^{\mathsf{SP}}_A(\mathcal{I}), \Gamma^{\mathsf{SP}}_X(\mathcal{I}))
\end{aligned}$$

Lemma 5.19 (Monotonicity)**.** Γ^{SP} *is monotone.*

Exploiting the monotonicity of the Jump Γ^{SP}, we can find fixed-points which yield nice interpretations for $\mathcal{L}_{\dot\epsilon}$. In particular, in these models every sentence $t \ \dot\epsilon\ \{x \mid \varphi\}$ will have the same value as $\varphi[t/x]$. Moreover, these models know that $\langle r, r\rangle$ is a paradoxical instance of $\dot\epsilon$. That is to say, in these models $r \ \dot\epsilon\ r$ has value $\mathfrak{p}$:

Theorem 5.20. *Define an interpretation* $\mathcal{I}^{\dot\epsilon}$ *to be* sound *if* $\mathcal{I}^{\dot\epsilon} \leq \Phi^{\mathsf{SP}}(\mathcal{I}^{\dot\epsilon})$*. Then*

(a) Γ^{SP} *has a least fixed-point* $\mathcal{I}_\mu$ *such that (i)* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \ \dot\epsilon\ \{x \mid \varphi\}) = \mathfrak{p}$ *for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$*, and (ii)* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \ \dot\epsilon\ \{x \mid \varphi\}) = \mathfrak{u}$ *for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{S}$.

(b) *Any fixed-point* $\mathcal{I}^{\dot\epsilon}_\kappa$ *of* Γ^{SP} *is such that* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \ \dot\epsilon\ \{x \mid \varphi\}) = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(\varphi[t/x])$*, for all* $\varphi \in \mathcal{L}_{\dot\epsilon}$*.*

(c) *Any fixed-point* $\mathcal{I}^{\dot\epsilon}_\kappa$ *of* Γ^{SP} *is such that* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \ \dot\epsilon\ \{x \mid \varphi\}) = \mathfrak{p}$ *for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$.

Proof. See Appendix. □

5.4 Generalizing to $\mathcal{L}_P$.

Generalizing from the above examples, one can extract a general method for obtaining a class of SP-fixed-point models for a language $\mathcal{L}_P$, for an arbitrary paradoxical predicate. So let P be a k-ary predicate; let $\mathbb{S}$ be the standard interpretation for $\mathcal{L}_P \backslash \{P\}$ containing codes for syntactic expressions of $\mathcal{L}_P$; let D be the set of codes of designated objects of P; let $ND := |\mathbb{S}| \backslash D$; let π be the principle we associate with P. Then, as a first step we can define as above the paradoxical and pathological instances of P:

Definition 5.21 (Paradoxical and pathological instances)**.** The class $\mathcal{P}$ of paradoxical instances of P contains all designated objects $t_1, \ldots, t_k$ of P such that

$$P(t_1, \ldots, t_k) \ \equiv^{\mathbb{S}}_{\mathsf{eq}}\ \neg\pi[t_1^{\mathbb{S}}, \ldots, t_k^{\mathbb{S}}]$$

[47] Again, we omit parameters for simplicity.

The class $\mathcal{U}$ of pathological but unparadoxical instances of P contains all designated objects $t_1, \ldots, t_k$ of P such that

$$P(t_1, \ldots, t_k) \;\equiv^{\mathbb{S}}_{\mathsf{eq}}\; \pi[t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k]$$

Definition 5.22 (P-Jump)**.** The *P-Jump*, Π^{SP}, is a function on pairwise disjoint triples $(E, A, X) = \mathcal{I}$ defined by:

$$\begin{aligned}
\Pi^{\mathsf{SP}}_E(\mathcal{I}) &:= \{\langle t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k\rangle \in D \mid \mathcal{V}_{\langle \mathbb{S},\mathcal{I}\rangle}(\pi[t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k]) = 1\}\\
\Pi^{\mathsf{SP}}_A(\mathcal{I}) &:= \{\langle t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k\rangle \in D \mid \mathcal{V}_{\langle \mathbb{S},\mathcal{I}\rangle}(\pi[t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k]) = 0\} \cup ND\\
\Pi^{\mathsf{SP}}_X(\mathcal{I}) &:= \{\langle t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k\rangle \in D \mid \mathcal{V}_{\langle \mathbb{S},\mathcal{I}\rangle}(\pi[t^{\mathbb{S}}_1, \ldots, t^{\mathbb{S}}_k]) = \mathfrak{p}\} \cup \mathcal{P}\\
\Pi^{\mathsf{SP}}(\mathcal{I}) &:= (\Pi^{\mathsf{SP}}_E(\mathcal{I}), \Pi^{\mathsf{SP}}_A(\mathcal{I}), \Pi^{\mathsf{SP}}_X(\mathcal{I}))
\end{aligned}$$

It is not difficult to verify that Π^{SP} is monotone, and that it has fixed-points, among which a least fixed-point differentiating between paradoxical and pathological instances of P. Also, applying this definition, one can define fixed-point models for $\mathcal{L}_{\mathrm{Ht}}$[48] within which: (i) the value of $\mathrm{Ht}(\ulcorner\varphi(x)\urcorner)$ is the same as $\neg\varphi(\ulcorner\varphi(x)\urcorner)$ for all $\varphi(x)$; (ii) $\#\mathrm{Ht}(x)$ is element of the paradox set interpreting Ht.

6 Analysis and comparison of models for truth

This section focuses on the truth predicate, and it compares the minimal fixed-points of the three semantics introduced above. The reason for concentrating on the minimal fixed-points is that, as suggested by [Kri75, p. 708], the minimal fixed-point is probably the most natural model for the intuitive concept of truth. The reason for concentrating on the truth predicate is that the majority of the investigations mentioned above also focuses on languages containing Tr. Nonetheless, even though the analysis will focus on models for $\mathcal{L}_{\mathrm{Tr}}$, similar remarks hold for an arbitrary $\mathcal{L}_P$.

As mentioned, the minimal fixed-point of Φ^{FDE} is not a dialetheic model, as $E \cap A = \varnothing$. But, as mentioned, here we are interested in those FDE models having Liars in $E \cap A$ and Truth-tellers outside $E \cup A$. In other words, we are interested in those fixed-points of Φ^{FDE} that are obtained by iterating the Jump on interpretations declaring Liars both true and false. Formally, this can be obtained by modifying Definition 3.2 of the FDE-Kripke Jump as follows. As in Definition 5.10, define Λ^{FDE} to be the set of sentences λ such that λ and $\neg\mathrm{Tr}\ulcorner\lambda\urcorner$ are equivalent in every standard FDE model $\langle\mathbb{N}, (E, A)\rangle$, such that $\#\varphi \in E$ (A) iff $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in E$ (A). Then, hoping that keeping the same notation for this Jump will not cause any confusion, define Φ^{FDE} as follows:

$$\begin{aligned}
\Phi^{\mathsf{FDE}}_E(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{FDE}}_{\langle\mathcal{N},(E,A)\rangle}(\varphi) \in \{1, \mathfrak{b}\}\} \cup \Lambda^{\mathsf{FDE}},\\
\Phi^{\mathsf{FDE}}_A(E, A) &:= \{\#\varphi \mid \mathcal{V}^{\mathsf{FDE}}_{\langle\mathcal{N},(E,A)\rangle}(\varphi) \in \{0, \mathfrak{b}\}\} \cup \Lambda^{\mathsf{FDE}} \cup \mathrm{NSt},\\
\Phi^{\mathsf{FDE}}(E, A) &:= (\Phi^{\mathsf{FDE}}_E(E, A), \Phi^{\mathsf{FDE}}_A(E, A)).
\end{aligned}$$

[48] Recall that Ht is the heterologicality predicate.

Notation. The minimal fixed-points of Φ^{SP}, Φ^{SK}, and Φ^{FDE} will be denoted by $\mathcal{I}_\mu^{\mathsf{SP}}$, $\mathcal{I}_\mu^{\mathsf{SK}}$, and $\mathcal{I}_\mu^{\mathsf{FDE}}$, respectively, or sometimes $(E_\mu^{\mathsf{SP}}, A_\mu^{\mathsf{SP}}, X_\mu^{\mathsf{SP}})$, $(E_\mu^{\mathsf{SK}}, A_\mu^{\mathsf{SK}})$, and $(E_\mu^{\mathsf{FDE}}, A_\mu^{\mathsf{FDE}})$. Additionally, for $L = \mathsf{SP}, \mathsf{SK}, \mathsf{FDE}$ and $\mathcal{I}_\kappa^L$ a fixed point, I write $\mathcal{V}_\mu^L(\varphi)$ instead of $\mathcal{V}_{\mathcal{I}_\mu^L}^L(\varphi)$.

The first question I am going to consider is what is the relationship between the minimal SK-fixed-point and the minimal FDE-fixed-point. The latter is sharper, in the sense that it can detect the difference between paradoxical and pathological sentences. But the disparity between them is deeper, as these models do not agree on what is *strictly true* and *strictly false*:

Observation 6.1. *(i)* $E_\mu^{\mathsf{SK}} \neq E_\mu^{\mathsf{FDE}} - A_\mu^{\mathsf{FDE}}$. *(ii)* $A_\mu^{\mathsf{SK}} \neq A_\mu^{\mathsf{FDE}} - E_\mu^{\mathsf{FDE}}$.

Proof. I provide two examples. Since Liars and Truth-tellers have value $\mathfrak{b}$ and $\mathfrak{u}$, respectively, in the minimal FDE fixed-point, it follows that the disjunction $\lambda \vee \tau$ has value 1 in the minimal FDE fixed-point. That is to say: $\mathcal{V}_\mu^{\mathsf{FDE}}(\lambda \vee \tau) = 1$, and hence $\#\lambda \vee \tau \in E_\mu^{\mathsf{FDE}} - A_\mu^{\mathsf{FDE}}$. However, $\lambda \vee \tau$ has value $\mathfrak{u}$ in the minimal SK fixed-point, and hence $\#\lambda \vee \tau \notin E_\mu^{\mathsf{SK}}$. Dually, $\mathcal{V}_\mu^{\mathsf{FDE}}(\lambda \wedge \tau) = 0 \neq \mathcal{V}_\mu^{\mathsf{SK}}(\lambda \wedge \tau) = \mathfrak{u}$ hence $\#\lambda \wedge \tau \in A_\mu^{\mathsf{FDE}} - E_\mu^{\mathsf{FDE}}$ but $\#\lambda \wedge \tau \notin A_\mu^{\mathsf{SK}}$. □

This observation shows that, *pace* Fitting, FDE-models do not preserve every insight about truth and falsity captured by SK-models.[49] As is well known, one attractive feature of the Kripkean minimal fixed-point is that what is true and what is false is *grounded* in non-semantic facts. What this means is that the truth and the falsity of any statement ultimately rests on whether certain sentences not involving Tr are true or false.[50] This is an important intuition about the intuitive concept of truth that the Kripkean model is able to capture. In §3.3 it was asked what is the price we would have to pay for the sharpness obtained with FDE models and it was pointed out that accepting contradictions could have been already a high price to pay. But Observation 6.1 shows that there is a different, and possibly higher, cost: there are sentences which are *strictly true* (*false*) in $\mathcal{I}_\mu^{\mathsf{FDE}}$, whose truth (falsity) is not grounded in non-semantic facts. Whether this is a high cost to pay or not, of course depends on one's view on groundedness. What is certain, though, is that if one accepts the paraconsistent semantics, then in order to obtain a sharper minimal fixed-point model, one has to give up not only consistency, but also the groundedness of truth. In other words, one can no longer claim that the truth and the falsity of any statement are grounded in non-semantic facts, and (more generally) one cannot preserve every insight about truth captured by SK-models. Also, it seems important to emphasize that advocates of FDE-models are faced with the following additional question: How can we philosophically justify that sentences like $\lambda \vee \tau$ are strictly true, or that sentences like $\lambda \wedge \tau$ are strictly false? Providing a plausible answer does not seem to be an easy task.

But what about the minimal fixed-points of the SK- and of the SP-Jump? Do they disagree on simple truths and simple falsities? As already hinted at when discussing Observation 5.6, it turns out that relationship between SK-models and SP-models is very different from that between

[49] Recall (see Introduction) that according to Fitting, FDE "loses none of the original insights" of SK-models.
[50] For more details on groundedness, see [Yab82], [Lei05], [Hal14, §17].

SK-models and FDE-models. Specifically, the minimal fixed-points of the SK- and of the SP-Jump fully agree on what is true and what is false, that is, the extension and the anti-extension of the two fixed-points coincide:[51]

Proposition 6.2. $(E^{\mathsf{SK}}_\mu, A^{\mathsf{SK}}_\mu) = (E^{\mathsf{SP}}_\mu, A^{\mathsf{SP}}_\mu)$.

Proof. Consider the two transfinite sequences leading to $\mathcal{I}^{\mathsf{SK}}_\mu$ and $\mathcal{I}^{\mathsf{SP}}_\mu$, i.e.:

Φ^{SK}-sequence

$$\begin{aligned}(E_0, A_0) &:= (\varnothing, \varnothing)\\ (E_{\alpha+1}, A_{\alpha+1}) &:= \Phi^{\mathsf{SK}}(E_\alpha, A_\alpha)\\ (E_\lambda, A_\lambda) &:= (\bigcup_{\alpha<\lambda} E_\alpha, \bigcup_{\alpha<\lambda} A_\alpha)\end{aligned}$$

Φ^{SP}-sequence

$$\begin{aligned}(E'_0, A'_0, X'_0,) &:= (\varnothing, \varnothing, \varnothing)\\ (E'_{\alpha+1}, A'_{\alpha+1}, X'_{\alpha+1}) &:= \Phi^{\mathsf{SP}}(E'_\alpha, A'_\alpha, X'_\alpha)\\ (E'_\lambda, A'_\lambda, X'_\lambda) &:= (\bigcup_{\alpha<\lambda} E'_\alpha, \bigcup_{\alpha<\lambda} A'_\alpha, \bigcup_{\alpha<\lambda} X'_\alpha)\end{aligned}$$

Due to Observation 5.6, $(E^{\mathsf{SK}}_\alpha, A^{\mathsf{SK}}_\alpha) = (E'^{\mathsf{SP}}_\alpha, A'^{\mathsf{SP}}_\alpha)$, for all α, hence in particular $(E^{\mathsf{SK}}_\mu, A^{\mathsf{SK}}_\mu) = (E'^{\mathsf{SP}}_\mu, A'^{\mathsf{SP}}_\mu)$. □

The *groundedness property* is preserved in the SP minimal fixed-point: only statements grounded in arithmetical facts are deemed true or false. More generally, Proposition 6.2 shows that every insight and every intuition about truth and falsity captured by $\mathcal{I}^{\mathsf{SK}}_\mu$ is preserved in $\mathcal{I}^{\mathsf{SP}}_\mu$. It shows that there is a very precise sense in which accepting $\mathcal{I}^{\mathsf{SP}}_\mu$ does not cause any loss. The minimal fixed-point $\mathcal{I}^{\mathsf{SP}}_\mu$, it may be said, is essentially a Kripkean model. But one can also show that, while the new model does not lose any insight about truth and falsity captured by $\mathcal{I}^{\mathsf{SK}}_\mu$, it also has some advantages. In particular, unlike $\mathcal{I}^{\mathsf{SK}}_\mu$, $\mathcal{I}^{\mathsf{SP}}_\mu$ is now able to detect (at least some) paradoxical instances of Tr, and to know the difference between them and pathological but unparadoxical ones.

But what are the paradoxical instances of Tr detected by $\mathcal{I}^{\mathsf{SP}}_\mu$? Do these coincide with sentences which are paradoxical in the sense of [Kri75]? The answer to the last question turns out to be negative. More precisely, every sentence which is paradoxical in $\mathcal{I}^{\mathsf{SP}}_\mu$ is also paradoxical *sensu* [Kri75], but the converse does not hold. To begin with, let me show that sentences which are paradoxical in $\mathcal{I}^{\mathsf{SP}}_\mu$ are undefined in every SK-fixed-point:

Lemma 6.3. *Let K be the set of sentences which are paradoxical in the sense of [Kri75], that is, put $K := \{\varphi \mid \varphi \notin \bigcup\{E \cup A \mid (E, A) = \Phi^{\mathsf{SK}}(E, A)\}\}$. Then $X^{\mathsf{SP}}_\mu \subseteq K$.*

Proof Sketch. Consider the transfinite sequence leading to $\mathcal{I}^{\mathsf{SP}}_\mu$, i.e.

$$\begin{aligned}(E_0, A_0, X_0,) &:= (\varnothing, \varnothing, \varnothing)\\ (E_{\alpha+1}, A_{\alpha+1}, X_{\alpha+1}) &:= \Phi^{\mathsf{SP}}(E_\alpha, A_\alpha, X_\alpha)\\ (E_\lambda, A_\lambda, X_\lambda) &:= (\bigcup_{\alpha<\lambda} E_\alpha, \bigcup_{\alpha<\lambda} A_\alpha, \bigcup_{\alpha<\lambda} X_\alpha)\end{aligned}$$

[51] As [Ros19, Proposition 4.29] noticed, this result does not hold for his canonical evaluation, which properly extends $\mathcal{I}^{\mathsf{SK}}_\mu$.

First one observes that, if $\varphi \in \Lambda$, then φ is undefined in every SK-fixed-point. Second, by primary induction on α and subinduction on φ, show that if $\varphi \in X_\alpha - \Lambda$, then φ is undefined in all fixed-points of Φ^{SK}. $\square$

Interestingly, though, the Kripkean paradoxical sentences do not coincide with the SP-paradoxical sentences:

Observation 6.4. $K \nsubseteq X^{\mathsf{SP}}_\mu$.

Proof. I provide an example. Let τ be a Truth-teller and λ a Liar. Then $(\tau \wedge \neg\tau) \vee \lambda \in K$ but $(\tau \wedge \neg\tau) \vee \lambda \notin X^{\mathsf{SP}}_\mu$, as $\mathcal{V}^{\mathsf{SP}}_\mu((\tau \wedge \neg\tau) \vee \lambda) = \mathfrak{u}$. $\square$

Remark 6.5. There is an intuitive way to generalize the previous observation: for any disjunction (conjunction) $\varphi \vee \psi$ ($\varphi \wedge \psi$) such that $\varphi \in K \cap X^{\mathsf{SP}}_\mu$ and ψ is an ungrounded sentence such that the only classical value it can receive is 0 (1), it holds that $\varphi \vee \psi$ ($\varphi \wedge \psi$) is element of K but not element of X^{SP}_μ. A more effective way of putting this is to say that, just as what is true or false is grounded in nonsemantic facts, what is paradoxical in $\mathcal{I}^{\mathsf{SP}}_\mu$ *depends* on Λ. This demarcates the present account of paradoxicality from that of [Kri75], where what is paradoxical also depends on logic. In fact, note that one reason why $(\tau \wedge \neg\tau) \vee \lambda \in K$ is that $\tau \wedge \neg\tau$ cannot obtain value 1. And this is a consequence of using the logic SK.[52]

The last observation may be seen as an inadequacy of the present construction, as it may be argued that every sentence in K is intuitively paradoxical, and one might suggest to substitute 'interestingly' with 'unfortunately' in the sentence introducing Observation 6.4. It has already been mentioned that intuitions about what it means for a sentence to be paradoxical are not as clear as intuitions about what it means for a sentence to be true or false. But it seems rather unplausible to claim that *every* sentence in K is intuitively paradoxical. To mention but a well-known example suggested by Gupta [Gup82], it can be observed that

$$\forall x \neg(\mathrm{Tr}(x) \wedge \neg\mathrm{Tr}(x)) \in K.$$

Intuitively, though, (or better: according to someone's intuitions) this law of classical logic is not paradoxical. So one might want a model within which this sentence is not paradoxical, but undefined. And indeed, the value of $\forall x \neg(\mathrm{Tr}(x) \wedge \neg\mathrm{Tr}(x))$ in the minimal fixed-point of the SP-Jump is actually $\mathfrak{u}$. Be that as it may, the claim here is neither that some sentences contained in K are not paradoxical, nor that every sentence in K is paradoxical. The only claim advocated here is that X^{SP}_μ singles out an interesting subclass of K—namely, the subclass given by (i) base paradoxical sentences and (ii) compositional sentences containing those as components.

Despite these important considerations about pre-theoretical intuitions on paradoxicality, it would certainly be desirable to have a natural fixed-point construction of K. Hence we leave the following question open for future research

[52] The idea of describing the set of paradoxical sentences as grounded in (or dependent on) Λ was suggested to me by Johannes Stern, whom I thank. For any inaccuracy in the Remark, I take full responsibility.

Open Question. *Is there a way to obtain a natural fixed-point construction of the triple* (E, A, X)*, where* (E, A) *is identical to the minimal fixed-point of* Φ^{SK} *and* $X = K$*?*

The present conjecture is that a supervaluational version of SP can serve as a basis for a positive answer.[53]

7 Concluding Remarks

The starting point of this article has been the question whether we can construct fixed-point models which have enough resources for distinguishing between different kinds of sentences and which can retain the insights about truth captured by SK-models, particularly the idea that truth is consistent and grounded in nonsemantic facts. We have seen that a positive answer to this question is possible. Specifically,

(i) like SK-models, but unlike FDE-models, SP-models do not satisfy any contradiction,

(ii) the minimal fixed point $\mathcal{I}_{\mu}^{\mathsf{SP}}$, like $\mathcal{I}_{\mu}^{\mathsf{SK}}$ but unlike $\mathcal{I}_{\mu}^{\mathsf{FDE}}$, is such that the only statements it deems true or false are those grounded in arithmetical facts;

(iii) like FDE-models, but unlike SK-models, SP-models are sharp enough to differentiate between paradoxical and pathological instances of P.

Before briefly commenting on revenge paradoxes, a methodological clarification is in order: the starting point of the above construction has *not* been the addition of the new truth-value Paradoxical. The starting point has been the addition of a third set into the interpretation of P-like predicates (see §4.1). The addition of a fourth truth-value, and the successive development of the semantics SP, has been a *consequence* of this new kind of semantic interpretation of a predicate. Of course, while I hope to have provided convincing philosophical reasons for the introduction of the truth-value $\mathfrak{p}$ and of the semantics SP, different semantic schemata could be defined to assign a value to $\mathcal{L}_P$-sentences into an SP-structure $\langle \mathcal{N}, (E, A, X)\rangle$.

7.1 Expressive weakness and revenge

A problem often associated with the account developed in [Kri75] is its expressive weakness and its non-immunity to revenge paradoxes.[54] The expressive weakness is due to the impossibility of expressing, in the object language, the semantic fact that undefined sentences are *not true*,[55] or, to use [Sch10]'s way of putting it, *something other than true*. The revenge paradox then follows:

[53] By 'supervaluational version of SP', I mean the following. Let SP^{-} be the three valued semantics obtained from SP by dropping the value $\mathfrak{u}$ (note that SP^{-} is nothing but SK with $\mathfrak{p}$ instead of $\mathfrak{u}$). Then a SP-supervaluation function $\mathcal{V}_{\mathcal{I}}^{\mathsf{SvSP}} : \mathrm{St}_{\mathcal{L}_{\mathrm{Tr}}} \longrightarrow \{1, 0, \mathfrak{p}, \mathfrak{u}\}$ can be defined as follow: for $v \in \{1, 0, \mathfrak{p}\}$, $\mathcal{V}_{\mathcal{I}}^{\mathsf{SvSP}}(\varphi) = v$ iff, for all total SP-interpretation $\mathcal{I}' \geq \mathcal{I}$ (see Df. 5.2) holds $\mathcal{V}_{\mathcal{I}'}^{\mathsf{SP}-}(\varphi) = v$; otherwise $\mathcal{V}_{\mathcal{I}}^{\mathsf{SvSP}}(\varphi) = \mathfrak{u}$.

[54] See for instance [Bur79].

[55] Note the difference between 'not true', without hyphen, and 'not-true', with hyphen. Undefined sentences, even though they are not 'not-true', they are 'not true'. Equivalently, one can say that the object language does not have a *strong* negation, but only a *weak* one. Cf. [Sch10].

if one extends the language with a predicate for the new semantic value $\mathfrak{u}$, one can construct so called *strengthened Liars* or *revenge sentences*, namely sentences saying of themselves that they are something other than true (i.e., not-true or undefined). Considerations of this kind have lead for instance Schlenker to claim that "Kripke's theory is only successful for a very small fragment of English". [Sch10, p. 376]

Let me emphasize that the above objection towards [Kri75] is *not* related to the expressive weakness of the object language, nor to revenge issues. The present account suffers from the same inadequacies: within our object language, we cannot express the semantic fact that a sentence has value paradoxical or undefined, and if we augmented the language with two predicates for the values 'paradoxical' and 'undefined', we could construct (via a suitable diagonal construction) revenge sentences saying of themselves that they are something other than true.[56]

The revenge of the Liar is a difficult issue, and the literature on it is possibly as vast as the literature on Liar paradox itself.[57] The fact that the construction above is prone to revenge might be considered a weakness. And that may well be so. However, it should also be emphasized that, when studying Liar-like paradoxes, there are different projects one can pursue. One of this consists in developing a sort of universal language with a type-free truth predicate, in which we can express every relevant notion that is expressible in natural languages.[58] But this is only one goal, certainly not the only one. Kripke, for example, doubts "that such a goal can be achieved" [Kri75, p. 714], and one might argue that the ghost of Tarski hierarchy will always be with us, no matter how we try to get rid of it. Yet, this does not mean that there is nothing more to say about Liar-like paradoxes. To name a recent example, [Lei19]'s goal "is not any kind of 'universal language' with a type-free truth-predicate, but merely to develop a model-theoretic theory of truth that does a bit better than Kripke's and some others' did" (p. 385). And *that* was the goal of this article, too.

Acknowledgements Thanks to Johannes Stern, Leon Horsten, Catrin Campbell-Moore, Graham Priest, Lorenzo Rossi, Lucas Rosenblatt, and an anonymous referee for very helpful comments on a draft of this paper. Thanks to Philip Welch, Francesco Paoli, and Adam Přenosil for useful discussions. I also wish to thank members of the Truth and Semantics reading group (Johannes Stern, Thomas Schindler, Poppy Maknowitz, Simone Picenni, Tedy Nenu, and Catrin Campbell-Moore) for a very insightful conversation. Special thanks to Karl-Georg Niebergall, with whom I have discussed the ideas contained in this paper when they were still vague and inaccurate. His comments, suggestions, and corrections led to a substantial improvement. This work was supported by the AHRC South, West and Wales Doctoral Training Partnership (SWW DTP), Grant No. AH/L503939/1-DTP1.

[56] Let me mention that in [Ros19]'s canonical evaluation, revenge sentences are assigned a set of equations with no solution in the value space.

[57] See [Bea07, Sha11, Sch13, MR20] for an overview.

[58] This hope is what underlies e.g. the view called *Embracing Revenge* [Coo07, Sch10, TTC16].

A Proofs

This appendix contains proofs of Theorem 5.15, Theorem 5.20, and Proposition 5.16. I begin with two auxiliary propositions.

Proposition A.1. *For all* $\varphi \in \mathcal{L}_{\mathrm{Tr}}, \mathrm{Tr}\ulcorner\varphi\urcorner \notin \Lambda$.

Proof. Let φ be arbitrary. To show that $\varphi \notin \Lambda$, it suffices to define an adequate interpretation $\langle \mathbb{N}, \mathcal{I}\rangle$ where $\mathrm{Tr}\ulcorner\varphi\urcorner$ and $\neg\mathrm{Tr}(\ulcorner\mathrm{Tr}(\ulcorner\varphi\urcorner)\urcorner)$ have a different value. This can be defined thus:

$$E := \{\#\mathrm{Tr}^{i}\ulcorner\varphi\urcorner \mid i \in \omega\}, A := \varnothing =: X,$$

where

$$\mathrm{Tr}^{i}(t) := \underbrace{\mathrm{Tr}\ldots\dot{\mathrm{Tr}}}_{n \text{ times}}(t),$$

for $\dot{\mathrm{Tr}}$ representing the primitive recursive function $\#t \mapsto \#\mathrm{Tr}(t)$. $\square$

Recall that, given a fixed-point $\mathcal{I}_\kappa^{\mathsf{SP}}$, I often adopt the abbreviation $\mathcal{V}_\kappa^{\mathsf{SP}}(\varphi) := \mathcal{V}_{\mathcal{I}_\kappa^{\mathsf{SP}}}^{\mathsf{SP}}(\varphi)$.

Proposition A.2. *Every fixed-point of* Φ^{SP} *is an adequate interpretation.*

Proof. Given a fixed-point $\mathcal{I}_\kappa^{\mathsf{SP}}$ of Φ^{SP}, one has to show that $\#\varphi \in Y_\kappa$ iff $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in Y_\kappa$, for $Y = E, A, X$. Suppose first $\#\varphi \in X_\kappa$. It follows that $\mathcal{V}_\kappa^{\mathsf{SP}}(\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathfrak{p}$, hence $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in \Phi_X^{\mathsf{SP}}(\mathcal{I}_\kappa) = X_\kappa$. Now suppose $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in X_\kappa$, which implies that either $\mathcal{V}_\kappa^{\mathsf{SP}}(\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathfrak{p}$, or that $\mathrm{Tr}\ulcorner\varphi\urcorner \in \Lambda$. Since by Proposition A.1 $\mathrm{Tr}\ulcorner\varphi\urcorner \notin \Lambda$, it follows that $\mathcal{V}_\kappa^{\mathsf{SP}}(\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathfrak{p}$. Hence, since $\mathcal{V}_\kappa^{\mathsf{SP}}(\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathfrak{p}$, we get $\#\varphi \in \mathcal{I}_\kappa$. With a similar argument, one shows that $\#\varphi \in Y_\kappa$ iff $\#\mathrm{Tr}\ulcorner\varphi\urcorner \in Y_\kappa$, for $Y = E, A$. $\square$

Theorem 5.15.

(a) Φ^{SP} *has a least fixed-point* $\mathcal{I}_\mu$ *which is such that* $\mathcal{V}_{\mathcal{I}_\mu}^{\mathsf{SP}}(\lambda) = \mathfrak{p}$ *for all* $\lambda \in \Lambda$ *and* $\mathcal{V}_{\mathcal{I}_\mu}^{\mathsf{SP}}(\tau) = \mathfrak{u}$ *for all* $\tau \in \Theta$.

(b) *Any fixed-point* $\mathcal{I}_\kappa$ *of* Φ^{SP} *is such that* $\mathcal{V}_{\mathcal{I}_\kappa}^{\mathsf{SP}}(\lambda) = \mathfrak{p}$ *for all* $\lambda \in \Lambda$.

(c) *Any fixed-point* $\mathcal{I}_\kappa$ *of* Φ^{SP} *is such that* $\mathcal{V}_{\mathcal{I}_\kappa}^{\mathsf{SP}}(\varphi) = \mathcal{V}_{\mathcal{I}_\kappa}^{\mathsf{SP}}(\mathrm{Tr}\ulcorner\varphi\urcorner)$, *for all* $\varphi \in \mathcal{L}_{\mathrm{Tr}}$.

Proof. I begin by showing that Φ^{SP} has a minimal fixed-point. Consider the sequences

Φ^{SK}-sequence

$$(E_0, A_0) := (\varnothing, \varnothing)$$
$$(E_{\alpha+1}, A_{\alpha+1}) := \Phi^{\mathsf{SK}}(E_\alpha, A_\alpha)$$
$$(E_\lambda, A_\lambda) := (\bigcup_{\alpha<\lambda} E_\alpha, \bigcup_{\alpha<\lambda} A_\alpha)$$

Φ^{SP}-sequence

$$(E'_0, A'_0, X'_0,) := (\varnothing, \varnothing, \varnothing)$$
$$(E'_{\alpha+1}, A'_{\alpha+1}, X'_{\alpha+1}) := \Phi^{\mathsf{SP}}(E'_\alpha, A'_\alpha, X'_\alpha)$$
$$(E'_\lambda, A'_\lambda, X'_\lambda) := (\bigcup_{\alpha<\lambda} E'_\alpha, \bigcup_{\alpha<\lambda} A'_\alpha, \bigcup_{\alpha<\lambda} X'_\alpha)$$

As remarked in the proof of Proposition 6.2, due to Observation 5.6, $(E_\alpha^{\mathsf{SK}}, A_\alpha^{\mathsf{SK}}) = (E_\alpha'^{\mathsf{SP}}, A_\alpha'^{\mathsf{SP}})$, for all α, hence in particular $(E_\mu^{\mathsf{SK}}, A_\mu^{\mathsf{SK}}) = (E_\mu'^{\mathsf{SP}}, A_\mu'^{\mathsf{SP}})$. What I have to show is that, at each α, $\mathcal{I}'_\alpha$ is an SP-interpretation, i.e., $X'_\alpha \cap (E'_\alpha \cup A'_\alpha) = \varnothing$. Suppose by i.h. that $\mathcal{I}'_\beta$ is an SP-interpretation. Consider an arbitrary $\varphi \in X'_{\beta+1} - \Lambda$. It follows that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(\varphi) = \mathfrak{p}$, hence $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(\varphi) \notin \{1, 0\}$, which is equivalent to $\#\varphi \notin E'_{\beta+1} \cup A'_{\beta+1}$. So let $\varphi \in X'_{\beta+1} \cap \Lambda$. Since $\varphi \in \Lambda$, by Lemma 6.3 we derive that $\varphi \notin E_\mu \cup A_\mu$. It follows that $\varphi \notin E_{\beta+1} \cup A_{\beta+1}$ hence $\varphi \notin E'_{\beta+1} \cup A'_{\beta+1}$. Since we have shown that the Φ^{SP}-sequence is an increasing sequence of SP-interpretations, by cardinality considerations it will reach a fixed-point, that can be shown to be contained in any other fixed-point of Φ^{SP}.[59]

I now show that $\mathcal{V}^{\mathsf{SP}}_\kappa(\lambda) = \mathfrak{p}$ for all $\lambda \in \Lambda$ and for all fixed-points $\mathcal{I}_\kappa$. So let $\varphi \in \Lambda$, which means that φ and $\neg\mathrm{Tr}\ulcorner\varphi\urcorner$ have the same value in every adequate interpretation. Since $\varphi \in \Lambda$, we have $\#\varphi \in \Phi^{\mathsf{SP}}_X(\mathcal{I}_\kappa) = X_\kappa$, which is equivalent to $\mathcal{V}^{\mathsf{SP}}_\kappa(\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathfrak{p} = \mathcal{V}^{\mathsf{SP}}_\kappa(\neg\mathrm{Tr}\ulcorner\varphi\urcorner)$. Since $\mathcal{I}_\kappa$ is adequate by Proposition A.2, we get $\mathcal{V}^{\mathsf{SP}}_\kappa(\neg\mathrm{Tr}\ulcorner\varphi\urcorner) = \mathcal{V}^{\mathsf{SP}}_\kappa(\varphi) = \mathfrak{p}$.

I next show that $\mathcal{V}^{\mathsf{SP}}_\mu(\tau) = \mathfrak{u}$ for all $\tau \in \Theta$. Let $\tau \in \Theta$ be arbitrary, and define

$$\mathcal{I}^{\mathsf{SK}} := \langle \{\#\mathrm{Tr}^i\ulcorner\tau\urcorner \mid i \in \omega\}, \varnothing\rangle \qquad \mathcal{I}^{\mathsf{SP}} := \langle \{\#\mathrm{Tr}^i\ulcorner\tau\urcorner \mid i \in \omega\}, \varnothing, \varnothing\rangle$$

$\mathcal{I}^{\mathsf{SP}}$ is adequate in the sense of Df. 5.9. Then consider the sequences

$$\begin{aligned} \mathcal{I}_0^{\mathsf{SK}} &:= \mathcal{I}^{\mathsf{SK}} & \qquad \mathcal{I}_0^{\mathsf{SP}} &:= \mathcal{I}^{\mathsf{SP}} \\ \mathcal{I}_{\alpha+1}^{\mathsf{SK}} &:= \Phi^{\mathsf{SK}}(\mathcal{I}_\alpha^{\mathsf{SK}}) & \qquad \mathcal{I}_{\alpha+1}^{\mathsf{SP}} &:= \Phi^{\mathsf{SP}}(\mathcal{I}_\alpha^{\mathsf{SP}}) \\ \mathcal{I}_\lambda^{\mathsf{SK}} &:= \bigcup_{\alpha<\lambda} \mathcal{I}_\alpha^{\mathsf{SK}} & \qquad \mathcal{I}_\lambda^{\mathsf{SP}} &:= \bigcup_{\alpha<\lambda} \mathcal{I}_\alpha^{\mathsf{SP}} \end{aligned}$$

By soundness of $\mathcal{I}_0^{\mathsf{SK}}$ and $\mathcal{I}_0^{\mathsf{SP}}$, we obtain fixed-points, say $\mathcal{I}_\xi^{\mathsf{SK}}$ and $\mathcal{I}_\xi^{\mathsf{SP}}$, identical modulo X_ξ^{SP}. Since $\mathcal{V}^{\mathsf{SK}}_{\mathcal{I}_\xi^{\mathsf{SK}}}(\tau) = 1 = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\xi^{\mathsf{SP}}}(\tau) = 1$, we obtain that $\tau \notin X_\mu^{\mathsf{SP}}$ by contraposition on Lemma 6.3.

The argument for (c) is standard, except that it uses that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(\lambda) = \mathfrak{p}$ for all $\lambda \in \Lambda$. □

Theorem 5.20.

(a) Γ^{SP} *has a least fixed-point* $\mathcal{I}_\mu$ *which is contained in every other. Moreover, (i) for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$, $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{p}$, *and (ii) for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{S}$ $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{u}$.

(b) *Any fixed-point* $\mathcal{I}_\kappa^{\dot\epsilon}$ *of* Γ^{SP} *is such that* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(\varphi[t/x])$, *for all* $\varphi \in \mathcal{L}_{\dot\epsilon}$.

(c) *Any fixed-point* $\mathcal{I}_\kappa^{\dot\epsilon}$ *of* Γ^{SP} *is such that* $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\kappa}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{p}$ *for all* $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$.

Proof. For (a), we show that every sound interpretation can be extended to a fixed-point containing it. Intuitively, the claim is that if we start the iteration of Γ^{SP} on a sound interpretation $\mathcal{I}^{\dot\epsilon}$, we end up with an SP-interpretation. To show this formally, consider the sequence:

$$\mathcal{I}_0^{\dot\epsilon} := \mathcal{I}^{\dot\epsilon}, \qquad \mathcal{I}_{\alpha+1}^{\dot\epsilon} := \Gamma^{\mathsf{SP}}(\mathcal{I}_\alpha^{\dot\epsilon}), \qquad \mathcal{I}_\lambda^{\dot\epsilon} := \bigcup_{\alpha<\lambda} \mathcal{I}_\alpha^{\dot\epsilon}.$$

[59]Recall that a sequence $(x_\alpha)_{\alpha\in On}$ is *strongly* increasing if for all α, $x_\alpha < x_{\alpha+1}$. The above sequence is only *weakly* increasing as for some β we have $\mathcal{I}_\beta^{\mathsf{SP}} = \mathcal{I}_{\beta+1}^{\mathsf{SP}}$ and the sequence remains constant from that point on.

By transfinite induction on α, we show that any $\mathcal{I}_\alpha$ is an SP-interpretation. The crucial observations is that, for all α,

$$\mathrm{R} \cap (E_\alpha \cup A_\alpha) = \varnothing, \tag{1}$$

To show (1), one can reason by transfinite induction. For $\alpha = 0$, suppose by reductio $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R} \cap (E_0 \cup A_0)$, for some $\langle t, \{x \mid \varphi\}\rangle$. Assume $\langle t, \{x \mid \varphi\}\rangle \in E_0$. Then $\mathcal{V}^{\mathsf{SP}}_0(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = 1$. By Df. 5.17 of R, we obtain $\mathcal{V}^{\mathsf{SP}}_0(\neg\varphi[t/x]) = 1$, which is the case iff $\mathcal{V}^{\mathsf{SP}}_0(\varphi[t/x]) = 0$. But then $\langle t, \{x \mid \varphi\}\rangle \in A_1$ and not in E_1, contradicting the soundness of $\mathcal{I}_0$. Symmetrically, it can be shown that $\langle t, \{x \mid \varphi\}\rangle \notin A_0$. For $\alpha = \beta + 1$, assume by i.h. that $\mathrm{R} \cap (E_\beta \cup A_\beta) = \varnothing$. Then, if $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$, then $\mathcal{V}_\beta(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) \notin \{1, 0\}$, hence $\langle t, \{x \mid \varphi\}\rangle \notin (E_{\beta+1} \cup A_{\beta+1})$. Since limits are just unions, (1) has been established.

The fact that for all α, $\mathcal{I}_\alpha$ is an SP interpretation then follows rather immediately. In fact, suppose that $\mathcal{I}_\beta$ is a disjoint triple, which implies that every sentence has a unique value. This immediately yields that $E_{\beta+1} \cap A_{\beta+1} = \varnothing$. Now suppose that $\langle t, \{x \mid \varphi\}\rangle \in X_{\beta+1}$. Then either $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$, which by (1) implies $\langle t, \{x \mid \varphi\}\rangle \notin E_{\beta+1} \cup A_{\beta+1}$, or $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(\varphi[t/x]) = \mathfrak{p}$, and hence $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(\varphi[t/x]) \notin \{1, 0\}$, and hence $\langle t, \{x \mid \varphi\}\rangle \notin E_{\beta+1} \cup A_{\beta+1}$.

By cardinality considerations, the weakly increasing sequence reaches a fixed-point.

The arguments for the remaining claims are similar to those of Theorem 5.15. I just show (c) and a-(ii), beginning with (c). Suppose $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{R}$. Then, for $\mathcal{I}_\kappa$ an arbitrary fixed-point of Γ^{SP}, we have $\langle t, \{x \mid \varphi\}\rangle \in \Gamma^{\mathsf{SP}}_X(\mathcal{I}_\kappa) = X_\kappa$, which is equivalent to $\mathcal{V}^{\mathsf{SP}}_\kappa(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{p}$ iff $\mathcal{V}^{\mathsf{SP}}_\kappa(\neg\varphi[t/x]) = \mathfrak{p} = \mathcal{V}^{\mathsf{SP}}_\kappa(\varphi[t/x])$.

As for (a)-(ii), we have to show that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\mu}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{u}$ for all $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{S}$. First notice that $\mathrm{S} \cap \mathrm{R} = \varnothing$. Then consider the sequence leading to the minimal fixed-point of Γ^{SP}, i.e., the one starting the iteration of Γ^{SP} on $\mathcal{I}_0 := \langle \varnothing, \varnothing, \varnothing\rangle$. Clearly, for any $\langle t, \{x \mid \varphi\}\rangle \in \mathrm{S}$, $\mathcal{V}^{\mathsf{SP}}_0(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{u} = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_0}(\varphi[t/v])$. Supposing by i.h. that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(t \mathrel{\dot\epsilon} \{x \mid \varphi\}) = \mathfrak{u} = \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}_\beta}(\varphi[t/v])$, we immediately get $\langle t, \{x \mid \varphi\}\rangle \notin \mathcal{I}^{\dot\epsilon}_{\beta+1}$. □

Let $\Phi^{\mathsf{SP}^\star}$ be defined just as Φ^{SP} but with $\Lambda^\star$ instead of Λ, and let $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$ be the minimal fixed-point of $\Phi^{\mathsf{SP}^\star}$. Then

Proposition 5.16. *(i) A Curry sentence κ is paradoxical in both $\mathcal{I}^{\mathsf{SP}}_\mu$ and $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$; (ii) the McGee sentence μ is paradoxical in $\mathcal{I}^{\mathsf{SP}}_\mu$ and undefined in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$; (iii) liar cycles are paradoxical in $\mathcal{I}^{\mathsf{SP}}_\mu$ and undefined in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$.*

Proof. (i) A Curry sentence (with a false consequent) is a sentence κ equivalent over $\mathbb{N}$ with $\mathrm{Tr}\ulcorner\kappa\urcorner \to \overline{0} = \overline{1}$, which in our framework is equivalent to $\neg\mathrm{Tr}\ulcorner\kappa\urcorner \vee \overline{0} = \overline{1}$.[60] Since $\neg\mathrm{Tr}\ulcorner\kappa\urcorner \vee \overline{0} = \overline{1}$ is equivalent over $\mathbb{N}$ to $\neg\mathrm{Tr}\ulcorner\kappa\urcorner$, it follows that $\kappa \in \Lambda^\star \subseteq \Lambda$, which in turn implies $\mathcal{V}^{\mathsf{SP}}_{\mu^{(\star)}}(\neg\mathrm{Tr}\ulcorner\kappa\urcorner) = \mathcal{V}^{\mathsf{SP}}_{\mu^{(\star)}}(\kappa) = \mathfrak{p}$.

[60] Note that, since $\mathcal{L}_{\mathrm{PA}}$ does not have a conditional, the analysis of Curry sentences amounts to an analysis of Liars with Boolean compounds. See [BGR20] for details.

(ii) For the McGee sentence,[61] let me begin by showing that μ is undefined in the minimal fixed-point constructed using $\Lambda^\star$. To begin with, μ is equivalent over $\mathbb{N}$ to

$$\exists x \neg \mathrm{Tr} \underset{\cdot}{f}(x, \ulcorner \mu \urcorner)),$$

where $\underset{\cdot}{f}(x,y)$ is a function symbol representing the primitive recursive operation

$$n, \#\varphi \mapsto \#\mathrm{Tr}^{n} \ulcorner \varphi \urcorner.$$

First one shows that $\mu \not\equiv^{\mathbb{N}}_{\mathsf{eq}} \neg \mathrm{Tr} \ulcorner \mu \urcorner$ by considering an interpretation like

$$\mathcal{I} = (E, A, X) := (\{\mu\}, \{\mathrm{Tr} \ulcorner \mu \urcorner\}, \emptyset),$$

and second one notices that by Proposition A.1, for $n > 0, \#\mathrm{Tr}^{n} \ulcorner \mu \urcorner \notin \Lambda^\star$. So for any $n, \mathrm{Tr}^{n} \ulcorner \mu \urcorner \notin \Lambda^\star$. Now if μ has value $\mathfrak{p}$ in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$, then, by Df. 5.3(e), for some $n_k \in \omega$, $\mathcal{V}^{\mathsf{SP}}_{\mu^\star}(\neg \mathrm{Tr} \underset{\cdot}{f}(\overline{n_k}, \ulcorner \mu \urcorner)) = \mathfrak{p}$, which is the case iff $\#\mathrm{Tr}^{n_k} \ulcorner \mu \urcorner \in X_{\mu^\star}$. Since we have just seen that $\mathrm{Tr}^{n_k} \ulcorner \mu \urcorner \notin \Lambda^\star$, it follows that $\mathcal{V}^{\mathsf{SP}}_{\alpha}(\mathrm{Tr}^{n_k} \ulcorner \mu \urcorner) = \mathfrak{p}$ for some level α in the construction of the minimal fixed-point. By transfinite induction on α, and subinduction on n, show that for α and for all $n, \mathcal{V}^{\mathsf{SP}}_{\alpha}(\mathrm{Tr}^{n} \ulcorner \mu \urcorner) \neq \mathfrak{p}$.

To show, to the contrary, that μ is paradoxical in $\mathcal{I}^{\mathsf{SP}}_{\mu}$, one shows that $\mu \in \Lambda$ as follows: suppose, for some adequate SP-interpretation $\mathcal{I}$, that $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\mu) = 1$. Then $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\exists x \neg \mathrm{Tr} \underset{\cdot}{f}(x, \ulcorner \mu \urcorner))) = 1$ iff, by Df. 5.3(e), for some $n \in \omega, \mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\neg \mathrm{Tr} \underset{\cdot}{f}(\overline{n}, \ulcorner \mu \urcorner))) = 1$ iff $\#\mathrm{Tr}^{n} \ulcorner \mu \urcorner \in A$. By adequacy, we have $\#\mathrm{Tr}^{i} \ulcorner \mu \urcorner \in A$ for all $i \in \omega$ hence in particular $i = 0$, hence $\mu \in A$ which is the case precisely if $\mathcal{V}^{\mathsf{SP}}_{\mathcal{I}}(\neg \mathrm{Tr} \ulcorner \mu \urcorner) = 1$. The same argument, *mutatis mutandis*, works for the other values $0, \mathfrak{p}, \mathfrak{u}$. Since $\#\mu \in \Lambda \subseteq X_{\mu}$, we have by adequacy of the minimal fixed-point that $\#\mathrm{Tr}^{i} \ulcorner \mu \urcorner \in X_{\mu}$ for all $i \in \omega$. Hence $\mathcal{V}^{\mathsf{SP}}_{\mu}(\neg \mathrm{Tr}(\overline{n}, \ulcorner \mu \urcorner)) = \mathfrak{p}$ for all $n \in \omega$, hence $\mathcal{V}^{\mathsf{SP}}_{\mu}(\exists x \neg \mathrm{Tr} \underset{\cdot}{f}(x, \ulcorner \mu \urcorner)) = \mathfrak{p} = \mathcal{V}^{\mathsf{SP}}_{\mu}(\mu)$.

(iii) We show the argument for a 2-Liar cycle, for n Liar cycles the argument being similar, beginning by showing that they are undefined in $\mathcal{I}^{\mathsf{SP}}_{\mu^\star}$. By diagonalization, find λ_1 and λ_2 such that

$$\begin{aligned} \lambda_1 &\equiv^{\mathbb{N}}_{\mathsf{eq}} \neg \mathrm{Tr} \ulcorner \lambda_2 \urcorner \\ \lambda_2 &\equiv^{\mathbb{N}}_{\mathsf{eq}} \mathrm{Tr} \ulcorner \lambda_1 \urcorner \end{aligned}$$

To see that $\lambda_i \not\equiv^{\mathbb{N}}_{\mathsf{eq}} \neg \mathrm{Tr} \ulcorner \lambda_i \urcorner$, for $i = 1, 2$ just consider the following interpretations

$$\langle \{\lambda_1\}, \{\lambda_2\}, \emptyset \rangle \qquad \langle \emptyset, \{\lambda_2, \lambda_1\}, \emptyset \rangle$$

Hence $\{\lambda_1, \lambda_2\} \cap \Lambda^\star = \varnothing$. It is then easy to verify that $\{\lambda_1, \lambda_2\} \cap \mathcal{I}_\alpha = \varnothing$ for all α leading to the minimal fixed-point of $\Phi^{\mathsf{SP}^\star}$.

In order to show that λ_1 and λ_2 are paradoxical in $\mathcal{I}^{\mathsf{SP}}_{\mu}$, one first shows that $\lambda_1 \in \Lambda$. It then follows that they both have value $\mathfrak{p}$ in $\mathcal{I}^{\mathsf{SP}}_{\mu}$. □

[61] It may be worth mentioning that in [Ros19], the McGee sentence has the same status as Liars and Curry sentences.

Remark A.3. Observe that λ_2 is not element of Λ. Its value is *contingent* on the value of λ_1.

References

[Bac13] Andrew Bacon. A new conditional for naive truth theory. *Notre Dame Journal of Formal Logic*, 54(1):87–104, 2013.

[Bea07] Jeffrey C Beall, editor. *Revenge of the Liar: New essays on the paradox.* Oxford University Press, 2007.

[Bea09] Jeffrey C Beall. *Spandrels of truth.* Oxford University Press, 2009.

[Bel77] ND Belnap. A useful four-valued logic. In *Modern uses of multiple-valued logic*, pages 5–37. Springer, 1977.

[Bel19] Nuel Belnap. How a computer should think. In Hitoshi Omori and Heinrich Wansing, editors, *New essays on Belnap-Dunn Logic*, pages 35–53. Springer, 2019.

[BGR20] Jc Beall, Michael Glanzberg, and David Ripley. Liar paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, fall 2020 edition, 2020.

[Bra71] Ross T Brady. The consistency of the axioms of abstraction and extensionality in a three-valued logic. *Notre Dame Journal of Formal Logic*, 12(4):447–453, 1971.

[Bur79] Tyler Burge. Semantical paradox. *The Journal of Philosophy*, 76(4):169–198, 1979.

[Bur86] John P Burgess. The truth is never simple. *Journal of Symbolic Logic*, pages 663–681, 1986.

[Can89] Andrea Cantini. Notes on formal theories of truth. *Mathematical Logic Quarterly*, 35(2):97–130, 1989.

[Can96] Andrea Cantini. *Logical frameworks for truth and abstraction: An axiomatic study*, volume 135. Elsevier, 1996.

[Can09] Andrea Cantini. Paradoxes, self-reference and truth in the 20th century. In Dov M. Gabbay and John Woods, editors, *Handbook of the history of logic*, volume 5, pages 875–1013. Elsevier, 2009.

[Cie02] Cezary Cieśliński. Heterologicality and incompleteness. *Mathematical Logic Quarterly*, 48(1):105–110, 2002.

[Coo07] Roy Cook. Embracing revenge: on the indefinite extensibility of language. In *Revenge of the liar: New essays on the paradox*, page 31. Oxford University Press, 2007.

[Dun76] J Michael Dunn. Intuitive semantics for first-degree entailments and ‘coupled trees’. *Philosophical Studies*, 29(3):149–168, 1976.

[Fef84] Solomon Feferman. Toward useful type-free theories. I. *The Journal of Symbolic Logic*, 49(1):75–111, 1984.

[Fie04] Hartry Field. The consistency of the naive theory of properties. *The Philosophical Quarterly*, 54(214):78–104, 2004.

[Fie08] Hartry Field. *Saving truth from paradox.* Oxford University Press, 2008.

[Fie14] Hartry Field. Naive truth and restricted quantification: Saving truth a whole lot better. *Review of Symbolic Logic*, 7(1), 2014.

[Fit48] Frederic B Fitch. An extension of basic logic. *The Journal of Symbolic Logic*, 13(2):95–106, 1948.

[Fit80] Frederic B Fitch. A consistent combinatory logic with an inverse to equality. *The Journal of Symbolic Logic*, 45(3):529–543, 1980.

[Fit86] Melvin Fitting. Notes on the mathematical aspects of Kripke’s theory of truth. *Notre Dame Journal of Formal Logic*, 27(1):75–88, 1986.

[Fit89] Melvin Fitting. Bilattices and the theory of truth. *Journal of Philosophical Logic*, 18(3):225–256, 1989.

[GB93] Anil Gupta and Nuel D Belnap. *The revision theory of truth.* Mit Press, 1993.

[Gil74] Paul C Gilmore. The consistency of partial set theory without extensionality. In *Axiomatic Set Theory, Proceedings of Symposia in Pure Mathematics*, volume 13, pages 147–153. American Mathematical Society, 1974.

[GN07] Kurt Grelling and Leonard Nelson. Bemerkungen zu den Paradoxien von Russell und Burali-Forti. In *Abhandlungen der Fries’schen Schule*, pages 301–334. Neue Folge, 1907.

[Gup82] Anil Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11(1):1–60, 1982.

[Hal14] Volker Halbach. *Axiomatic theories of truth.* Cambridge University Press, second edition, 2014.

[Her82a] Hans Herzberger. Naive semantics and the liar paradox. *The Journal of Philosophy*, 79(9):479–497, 1982.

[Her82b] Hans G Herzberger. Notes on naive semantics. *Journal of Philosophical Logic*, 11(1):61–102, 1982.

[Kle71] Stephen C Kleene. *Introduction to metamathematics.* North Holland, Amsterdam, 1971.

[Kre09] Philip Kremer. Comparing fixed-point and revision theories of truth. *Journal of Philosophical Logic*, 38(4):363–403, 2009.

[Kri75] Saul Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716, 1975.

[Lei99] Hannes Leitgeb. Truth and the Liar in De Morgan-valued models. *Notre Dame Journal of Formal Logic*, 40(4):496–514, 1999.

[Lei05] Hannes Leitgeb. What truth depends on. *Journal of Philosophical Logic*, 34(2):155–192, 2005.

[Lei19] Hannes Leitgeb. HYPE: A system of hyperintensional logic (with an application to semantic paradoxes). *Journal of Philosophical Logic*, 48(2):305–405, 2019.

[Mad83] Penelope Maddy. Proper classes. *The Journal of Symbolic Logic*, 48(1):113–139, 1983.

[MR20] Julien Murzi and Lorenzo Rossi. Generalized revenge. *Australasian Journal of Philosophy*, 98(1):153–177, 2020.

[MW75] Robert L Martin and Peter W Woodruff. On representing 'true-in-L' in L. *Philosophia*, 5(3):213–217, 1975.

[PBAG06] Graham Priest, Jeffrey C Beall, and Bradley Armour-Garb, editors. *The law of non-contradiction: New philosophical essays*. Clarendon Press, 2006.

[Pri94] Graham Priest. The structure of the paradoxes of self-reference. *Mind*, 103(409):25–34, 1994.

[Pri02] Graham Priest. Paraconsistent logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of philosophical logic*, volume 5, pages 287–393. Dordrecht: Kluwer Academic Publishers, second edition, 2002.

[Pri06] Graham Priest. *In contradiction*. Oxford University Press, 2006.

[Res17] Greg Restall. Fixed-point models for theories of properties and classes. *The Australasian Journal of Logic*, 14(1):226–241, 2017.

[Rip15] David Ripley. Paraconsistent logic. *Journal of Philosophical Logic*, 44(6):771–780, 2015.

[Ros16] Lorenzo Rossi. Adding a conditional to Kripke's theory of truth. *Journal of Philosophical Logic*, 45(5):485–529, 2016.

[Ros19] Lorenzo Rossi. A unified theory of truth and paradox. *The Review of Symbolic Logic*, 12(2):209–254, 2019.

[RW07] Agustın Rayo and Philip Welch. Field on revenge. In JC Beall, editor, *Revenge of the Liar*, pages 234–249. Oxford University Press, 2007.

[Sch10] Philippe Schlenker. Super liars. *The Review of Symbolic Logic*, 3(3):374–414, 2010.

[Sch13] Kevin Scharp. *Replacing truth.* OUP Oxford, 2013.

[Sha11] Lionel Shapiro. Expressibility and the liar's revenge. *Australasian Journal of Philosophy*, 89(2):297–314, 2011.

[Soa99] Scott Soames. *Understanding truth.* Oxford University Press on Demand, 1999.

[TTC16] Nicholas Tourville and Roy T Cook. Embracing the technicalities: Expressive completeness and revenge. *The Review of Symbolic Logic*, 9(2):325–358, 2016.

[Vis84] Albert Visser. Four valued semantics and the Liar. *Journal of Philosophical Logic*, 13(2):181–212, 1984.

[Wel01] Philip D Welch. On Gupta-Belnap revision theories of truth, Kripkean fixed points, and the next stable set. *Bulletin of Symbolic Logic*, pages 345–360, 2001.

[Wel11] Philip D Welch. Truth, logical validity and determinateness: A commentary on Field's Saving Truth from Paradox. *Rev. Symb. Log.*, 4(3):348–359, 2011.

[Wel14] Philip D Welch. Some observations on truth hierarchies. *Rev. Symb. Log.*, 7(1):1–30, 2014.

[Woo84] Peter W Woodruff. Paradox, truth and logic part I: Paradox and truth. *Journal of Philosophical Logic*, 13(2):213–232, 1984.

[Yab82] Stephen Yablo. Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11(1):117–137, 1982.