Synthese (2022) 200:71
https://doi.org/10.1007/s11229-022-03625-x

ORIGINAL RESEARCH

Paradoxicality in Kripke's theory of truth

Lucas Rosenblatt[1] · **Camila Gallovich**[2]

Received: 12 August 2021 / Accepted: 17 February 2022 / Published online: 5 March 2022

Abstract
A lot has been written on solutions to the semantic paradoxes, but very little on the topic of general theories of paradoxicality. The reason for this, we believe, is that it is not easy to disentangle a solution to the paradoxes from a specific conception of what those paradoxes consist in. This paper goes some way towards remedying this situation. We first address the question of what one should expect from an account of paradoxicality. We then present one conception of paradoxicality that has been offered in the literature: the fixed-point conception. According to this conception, a statement is paradoxical if it cannot obtain a classical truth-value at any fixed-point model. In order to assess this proposal rigorously we provide a non-metalinguistic characterization of paradoxicality and we evaluate whether the resulting account satisfies a number of reasonable desiderata.

Keywords Fixed-point semantics · Non-classical logic · Truth · Semantic paradoxes

1 Introduction

A lot has been written on solutions to the semantic paradoxes, but very little on the topic of general accounts of semantic paradoxicality. The reason for this, we believe, is that it is not easy to disentangle a solution to these paradoxes from a specific conception of what they consist in. Here we will try to do things differently. We will focus on a conception of semantic paradoxicality that is (or that at least aims to be) independent to some extent of specific solutions to the paradoxes. In other words, what we are

T.C. : Non-Classical Approaches to Paradox, Lead Guest Editor : Julien Murzi, Rossi Lorenzo, Topey Brett.

✉ Lucas Rosenblatt
lucasrosenblatt83@gmail.com

Camila Gallovich
camila.gallovich@gmail.com

1 National Council of Scientific and Technical Research–University of Buenos Aires, Bulnes 642, 1176 Buenos Aires, Argentina

2 University of Buenos Aires, Puan 480, Buenos Aires, Argentina

primarily interested in is not a theory of truth that intends to provide a solution to the liar paradox and other semantic antinomies, but a conception of paradoxicality that seeks to explain what makes a statement paradoxical. As Anil Gupta once suggested:

> "[T]he behavior of paradoxes (...) is so similar across different logics and semantics that it is fair to demand that any account of them be general, that it apply uniformly across the whole range of logics and semantics." (Gupta 2005, p. 143).

A number of different conceptions of semantic paradoxicality have been overtly or tacitly endorsed in the literature.[1] Our modest aim here is to analyze just one conception inspired by Kripke's paper on truth, and to focus on theories that fall within its purview.[2] Arguably, the fundamental insight of Kripke's paper is what Michael Kremer (1988, p. 228) once called "the fixed-point conception of truth". This is the idea that the meaning of the truth predicate is given by the claim that the circumstances under which one may assert of a statement that it is true (false) are exactly the same as the circumstances under which one may assert (deny) that statement. This can be made precise through the use of fixed-points models. A fixed-point (in Kripke's sense) is a model such that for every statement ϕ, the truth-value of ϕ is the same as the truth-value of the statement asserting that ϕ is true. The main thought underpinning the accounts of paradoxicality we will consider is what we shall call "the fixed-point conception of paradoxicality".[3] According to this, a statement is paradoxical if, and only if, there is no fixed-point at which it obtains a classical truth-value, where 'classical truth-value' is meant to exclude statements that are neither-true-nor-false and statements that are both-true-and-false.[4,5]

[1] Just to give the reader an idea, we can identify (i) the naive conception of paradoxicality (see Cook (2011) and Hsiung (2021)); (ii) the conception of paradoxicality as non-normalizability (cf. Prawitz (1965) and Tennant (1982)); (iii) the revision-theoretic conception (cf. Gupta (1982)); (iv) the inclosure-based conception (see Priest (1994)); and (v) the graph-theoretic conception (cf. Walicky (2017) and Rossi (2019), for a couple of recent examples). Some of these conceptions intend to cover all sorts of paradoxes. For the purposes of this paper we are only interested in *semantic* paradoxicality, and we are thus ignoring paradoxes that affect non-semantic concepts.

[2] Cf. Kripke (1975). Although probably Kripke would not subscribe to some of the ideas that we will put forward below—specially to the view that paradoxicality can behave non-classically in certain circumstances. More recently, the fixed-point conception has been discussed in Cook and Tourville (2020), Cook and Tourville (2016), Cook (2020), Castaldo (2021), Rosenblatt (2021) and Gallovich and Rosenblatt (2022).

[3] Our use of the definite description '*the* fixed-point conception of paradoxicality' should be taken with some caution. There is a sense in which the use of fixed-points is pervasive. For example, most (if not all) of the conceptions mentioned in Footnote 1 can probably be defined in terms of (the non-existence of) fixed-points. Kripke's construction is only a very specific example of the general applicability of fixed-points. However, the Kripke-inspired conception we will discuss explicitly relies on special semantic structures called "fixed-point *models*" (on which more shortly), and it does so in a very direct and blatant way. Thus, our more restrictive use of the term 'fixed-point' should not mislead. Thanks to an anonymous reviewer for urging us to clarify this.

[4] Traditionally, paradoxicality is thought to be a property of arguments, and a statement is said to be paradoxical only in a derivative sense—a statement is paradoxical because it contributes to the generation of paradoxical arguments. However, given that in the fixed-point conception paradoxicality is typically attached to statements, we will assume that it is statements that are the (primary) bearers of paradoxicality. This may be contentious, but a discussion would be beyond the scope of the paper. We take it up in ongoing work.

[5] We think that ultimately our account of paradoxicality should apply to natural languages, so we talk about paradoxical *statements*. A statement, as we are understanding it, is a declarative meaningful

In order to assess theories of semantic paradoxicality, we need to consider a number of rather natural desiderata that, we think, apply to such accounts. Some of them consist in the possession of certain theoretical virtues like consistency, naturalness, simplicity, explanatory power, etc., that play a role in any context where one is engaged in theory-choice. But there are other desiderata that are specific to theories of paradoxicality. First, it seems reasonable to say that a theory of paradoxicality ought to offer an explanation of what makes a statement paradoxical—that is, one should be in a position to identify some property such that a statement is paradoxical if and only if it has that property. Second, the theory should sanction a number of principles connecting the notion of paradoxicality with other notions, like negation, conjunction, truth and so on. For example, it should arguably validate the inference from the claim that some statement is paradoxical to the claim that the negation of that statement is paradoxical, among other things. Third, the theory ought to treat potential revenge paradoxes involving the notion of paradoxicality in roughly the same manner as it treats the truth-theoretic paradoxes. In other words, the theory must offer a unified treatment of 'ordinary' paradoxes and revenge paradoxes. Fourth, the theory should agree with our intuitions about paradoxicality in a large number of cases. In particular, it ought to establish the paradoxicality of statements that are typically viewed as semantic paradoxes, like the liar and its ilk.[6]

The main goal of the paper is to analyze theories that are based on the fixed-point conception and to ascertain how well they score on the desiderata that we have just proposed. In order to do this, we build on previous work by one of us (Rosenblatt 2021). The point of that paper was to respond to an objection posed by Julien Murzi and Lorenzo Rossi to non-classical accounts of truth (Murzi and Rossi 2020). The authors suggest that these accounts breed revenge paradoxes when they are coupled with the thought that classical reasoning can be recaptured in unparadoxical circumstances. In Rosenblatt (2021) it is shown that non-classical theorists can represent the concept of paradoxicality without falling prey to such revenge paradoxes by providing a formal fixed-point semantics for a language extended with a paradoxicality predicate. In this paper we attempt to offer a conceptual justification of the conception of paradoxicality underpinning that fixed-point semantics. To the best of our knowledge, there is next to nothing written on how to assess and evaluate approaches that directly offer an account of paradoxicality, as opposed to a mere solution to the paradoxes. Thus, the paper can be seen as taking the first steps towards a careful study of such accounts by analyzing one specific case, which uses fixed-point models in the explanation of paradoxes.

Footnote 5 continued
(non-ambiguous) type sentence together with a possible context of utterance. Of course, since in this paper our goal will be to characterize paradoxicality for a formal language, this will not be too important, and in fact it will be harmless to use 'statement' and 'sentence' (or even 'formula') interchangeably. The only exception to this occurs in Sect. 3.5, where we consider sentences that fail to express a proposition.

6 The list is not meant to be exhaustive. We are only suggesting that in evaluating and comparing different accounts of paradoxicality one should bear these desiderata in mind. For one thing, there are other general desiderata that play a role in theory-choice in science and that we have not even mentioned, such as predictive power, unificatory power, fertility, etc. For another, we could expand the list—following Hanness Leitgeb (2007)—by importing some criteria that play a role in the case of theories of truth. For example, we think that it is reasonable to require a theory of paradoxicality to be couched in a language that is rich enough to code facts about its own syntax. We also think that the paradoxicality predicate ought to be untyped.

The rest of the paper is structured as follows. We start by providing a non-metalinguistic characterization of a Kripke-inspired notion of paradoxicality based on the fixed-point semantics offered in Rosenblatt (2021) (Sect. 2). After that, we examine whether the account satisfies each of the desiderata one-by-one and we discuss one aspect of the fixed-point conception that we view as a virtue, its flexibility (Sect. 3). We then consider a potential limitation of this approach (Sect. 4). To finish, we offer some concluding remarks (Sect. 5).

2 Fixed-point semantics for the paradoxicality predicate

In his *Outline of a Theory of Truth* Kripke showed that one can start from a classical interpretation for a first-order base language without a truth predicate, and then construct various partial interpretations for the language containing the predicate. These interpretations—called "fixed-points"—are such that for every statement ϕ, the truth-value of ϕ is the same as the truth-value of the statement asserting that ϕ is true. In Kripke's approach a paradoxical statement is defined as a statement that does not obtain a classical truth-value at any fixed-point.[7]

Kripke's definition of paradoxicality is given in a set-theoretic metalanguage that obeys classical logic, not in the object language itself. He thinks this is unproblematic. In a footnote immediately after his celebrated phrase on the ghost of the Tarski hierarchy, Kripke (1975, p. 714) wrote that "[s]uch semantical notions as "grounded", "paradoxical", etc. belong to the metalanguage. This situation seems to me to be intuitively acceptable; in contrast to the notion of truth, none of these notions is to be found in natural language in its pristine purity (...)". Pace Kripke, we believe that this aspect of the original fixed-point approach might be viewed as a significant shortcoming.

First, one could be skeptical about the possibility of drawing a sharp distinction between notions that, as Kripke puts it, can be found in natural language in its pristine purity, and notions that cannot. It seems to us that this is more of a spectrum and that paradoxicality is a self-applicable predicate in roughly the same way that truth is. For instance, if someone asserts that not all statements are paradoxical, this assertion, taken at face value, can be instantiated not only by statements like $0 = 0$ or 'Paris is the capital of Brazil', but also by statements involving the paradoxicality predicate, like 'the liar is paradoxical' or the very statement 'not all statements are paradoxical'. Moreover, it is not hard to think of Nixon-like cases of assertions involving applications of the predicate 'is paradoxical' wherein it is not possible to assign types (or 'levels') without altering the intended meaning of the assertions. We can say, tweaking Kripke's words, that any statement, even those featuring the paradoxicality predicate, should be allowed to seek its own level.[8] Hence, we expect our theory of paradoxicality to be connected to some extent to our use of the paradoxicality predicate in natural language.[9]

7 Precursors of the use of fixed-points in the analysis of paradoxes include Lawvere (1969), Martin and Woodruff (1975) and Gilmore (1974), among others.

8 Cf. Kripke (1975, p. 696) for the original phrase and for his well-known diagnosis of the Nixon example.

9 Thanks are due to Luca Castaldo for discussion of this point.

Secondly, the notion of paradoxicality plays an important explanatory role in Kripke's theory. Thus, it is reasonable to expect an object-language treatment of it. This is so, we submit, even if one is not seeking to construct a "universal language" capable of expressing every intelligible semantic notion. If one is not in a position to offer an adequate object-language treatment of the notion of paradoxicality (or any notion that plays an important explanatory role in one's theory), then there are reasons to doubt the overall coherence of the theory.

To provide an object-language treatment of paradoxicality we can use the fixed-point semantics offered in Rosenblatt (2021).[10] The idea is to start from one of Kripke's interpretations for the truth predicate and then construct a different interpretation for the language containing the paradoxicality predicate. The new interpretation will be such that for any statement ϕ that is paradoxical in Kripke's sense, one is licensed to assert in the object-language that ϕ is paradoxical. To make this idea precise, we need a modicum of formal machinery. As our background theory of syntax we will rely on Peano arithmetic, PA. We will use $\mathcal{L}_{PA}$ for the language of PA with its usual signature $\{0, s, +, \times\}$, and $\mathcal{L}_+$ for the language that results from $\mathcal{L}_{PA}$ by adding a truth predicate, $Tr(x)$, and a predicate, $Par(x)$, standing for the notion of paradoxicality.[11] We assume a fixed canonical Gödel numbering for $\mathcal{L}_+$-expressions and we follow the usual practice of writing $\ulcorner \phi \urcorner$ for the Gödel code of the statement ϕ.

A model $\mathcal{M}$ for $\mathcal{L}_+$ is a structure $\langle \mathbb{N}, (E_\mathcal{T}, A_\mathcal{T}), (E_\mathcal{P}, A_\mathcal{P}) \rangle$, where $\mathbb{N}$ is the standard model of $\mathcal{L}_{PA}$. The other two components of $\mathcal{M}$, $(E_\mathcal{T}, A_\mathcal{T})$ and $(E_\mathcal{P}, A_\mathcal{P})$, are pairs of subsets of $|\mathbb{N}|$, the domain of $\mathbb{N}$. The first pair, $(E_\mathcal{T}, A_\mathcal{T})$, interprets $Tr(x)$. $E_\mathcal{T}$ stands for the extension of $Tr(x)$ and $A_\mathcal{T}$ stands for its antiextension. The extension, $E_\mathcal{T}$, is the set of (codes of) statements that are true at the model, and the anti-extension, $A_\mathcal{T}$, is the set of (codes of) statements that are false at the model (or codes that do not stand for statements). The second pair, $(E_\mathcal{P}, A_\mathcal{P})$, interprets the paradoxicality predicate, $Par(x)$. Thus, $E_\mathcal{P}$ is the set of (codes of) statements that are paradoxical in the model and $A_\mathcal{P}$ is the set of (codes of) statements that are not paradoxical in the model (which includes codes that do not stand for statements). Interpretations of this kind leave room for gaps, that is, (codes of) statements that are neither in the extension nor in the anti-extension of the corresponding predicate. As a result, the construction we are about to put forward yields a partial interpretation of truth and paradoxicality.

To assign truth-values to the statements of $\mathcal{L}_+$ we will rely on the three-valued strong Kleene schema.[12] A valuation $v_\mathcal{M}$ based on a model $\mathcal{M}$ is a function from the statements of $\mathcal{L}_+$ to the set of semantic values $\{1, \frac{1}{2}, 0\}$ satisfying the following conditions:

- $v_\mathcal{M}(s = t) = 1$ if and only if $s^\mathcal{M} = t^\mathcal{M}$; $v_\mathcal{M}(s = t) = 0$, otherwise.
- $v_\mathcal{M}(\neg\phi) = 1 - v_\mathcal{M}(\phi)$.
- $v_\mathcal{M}(\phi \wedge \psi) = min(v_\mathcal{M}(\phi), v_\mathcal{M}(\psi))$.

10 Cf. Rosenblatt and Szmuc (2014) for a similar kind of model-theoretic construction.

11 It would be possible to emulate the paradoxicality predicate $Par(x)$ using $Tr(x)$ together with a paradoxicality operator, $\mathcal{O}^\mathcal{P}$. That is, $Par(x)$ can be explicitly defined as $\mathcal{O}^\mathcal{P} Tr(x)$. So our choice of employing a predicate rather than an operator for paradoxicality is purely conventional.

12 We are focusing on this schema just for definiteness, but we are not committing ourselves to it. In fact, part of the appeal of the fixed-point conception is that it is compatible with other schemata as well. We will come back to this below.

- $v_{\mathcal{M}}(\forall x\phi) = min\{v_{\mathcal{M}'}(\phi) : \mathcal{M}' \text{ is an } x\text{-variant of } \mathcal{M}'\}$.[13]

If t is a $\mathcal{L}_{+}$-term, we will use the notation $t^{\mathcal{M}}$ for the denotation of t in the model $\mathcal{M}$. We say that for any model $\mathcal{M}$: $t^M \in E_{\mathcal{T}}$ if and only if $t^M = \ulcorner\phi\urcorner$ and $v_{\mathcal{M}}(\phi) = 1$; and $t^M \in A_{\mathcal{T}}$ if and only if (i) $t^M = \ulcorner\phi\urcorner$ and $v_{\mathcal{M}}(\phi) = 0$, or (ii) $t^{\mathcal{M}}$ is not the code of a statement. Then the semantic clause for statements of the form $Tr(t)$ can be given as follows:

- $$v_{\mathcal{M}}(Tr(t)) = \begin{cases} 1 & \text{if } t^M = \ulcorner\phi\urcorner \text{ and } v_{\mathcal{M}}(\phi) = 1 \\ 0 & \text{if } t^M = \ulcorner\phi\urcorner \text{ and } v_{\mathcal{M}}(\phi) = 0, \\ & \quad \text{or } t^{\mathcal{M}} \text{ is not the code of a statement} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

The crucial property of these valuations is that for any statement ϕ, it is the case that $v_{\mathcal{M}}(\phi) = v_{\mathcal{M}}(Tr\ulcorner\phi\urcorner)$. If a model $\mathcal{M}$ is such that the valuation $v_{\mathcal{M}}$ has this property, we will say that $\mathcal{M}$ is a *Kripke fixed-point*. Intuitively, Kripke fixed-points vindicate the thought that the circumstances under which one may assert of a statement that it is true (false) are exactly the same as the circumstances under which one may assert (deny) that statement.

We haven't yet explained how to interpret the paradoxicality predicate. According to Kripke's account, when one asserts that some statement is paradoxical, one is making a claim about its behaviour across different fixed-points. One may assert of a statement that it is paradoxical if, and only if, there are no circumstances under which one may assert that statement and there are no circumstances under which one may assert its negation. This gives the fixed-point conception a modal flavor. To evaluate attributions of paradoxicality at some fixed-point one must take into account fixed-points different from it. In particular, since a paradoxical statement is one that does not obtain a classical truth-value at any fixed-point, the paradoxicality predicate can be very naturally seen as a modal predicate of sorts, one that tracks down how statements behave at different fixed-points. Consequently, we need an additional definition to interpret $Par(x)$.

Definition (*Extension*) Let $\mathcal{M}$ be $\langle \mathbb{N}, (E_{\mathcal{T}}, A_{\mathcal{T}}), (E_{\mathcal{P}}, A_{\mathcal{P}})\rangle$ and let $\mathcal{M}'$ be $\langle \mathbb{N}, (E'_{\mathcal{T}}, A'_{\mathcal{T}}), (E'_{\mathcal{P}}, A'_{\mathcal{P}})\rangle$. We will say that $\mathcal{M}'$ extends $\mathcal{M}$ (in notation, $\mathcal{M} \preccurlyeq \mathcal{M}'$) if and only if the following four inclusions hold: (i) $E_{\mathcal{T}} \subseteq E'_{\mathcal{T}}$, (ii) $A_{\mathcal{T}} \subseteq A'_{\mathcal{T}}$, (iii) $E_{\mathcal{P}} \subseteq E'_{\mathcal{P}}$ and (iv) $A_{\mathcal{P}} \subseteq A'_{\mathcal{P}}$.

With this notion at our disposal we can now say that the truth-value of a statement $Par(\ulcorner\phi\urcorner)$ at a model depends on how ϕ behaves at the different fixed-points *extending* that model. The account resembles a possible world semantics in that the fixed-points play the role of possible worlds and the notion of extension plays the role of the accessibility relation. Thus, a statement $Par\ulcorner\phi\urcorner$ is true at a model if and only if ϕ is $\frac{1}{2}$ at every fixed-point extending that model.

[13] Of course, other logical expressions, such as $\vee$ and $\exists$, can be defined in terms of these. Also, *min* stands for the minimum operation and an *x-variant* of a model $\mathcal{M}$ is a model that is exactly like $\mathcal{M}$ except perhaps in what it assigns to the variable x. To simplify things, we leave the assignment function (which assigns objects in $|\mathbb{N}|$ to the variables of $\mathcal{L}_{+}$) implicit in the presentation of the models.

Every model $\mathcal{M}$ has an associated set whose members are all the Kripke fixed-points that extend it—we call this set $\mathcal{M}^{ext}$. More precisely, $\mathcal{M}^{ext} = \{\mathcal{M}' : \mathcal{M}'$ is a Kripke fixed-point and $\mathcal{M} \preccurlyeq \mathcal{M}'\}$. Then, one can interpret $Par(\ulcorner\phi\urcorner)$ at an specific interpretation $v_{\mathcal{M}}$ by looking at the behaviour of ϕ across the set of Kripke fixed-points that extend $\mathcal{M}$, i.e. across the members of $\mathcal{M}^{ext}$.

- $$v_{\mathcal{M}}(Par(t)) = \begin{cases} 1 & \text{if } t^{\mathcal{M}} = \ulcorner\phi\urcorner \text{ and } \forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2} \\ 0 & \text{if } t^{\mathcal{M}} = \ulcorner\phi\urcorner, \text{ and } \forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = 1 \\ & \quad \text{or } \forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = 0; \\ & \text{or } t^{\mathcal{M}} \text{ is not the code of a statement} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

If a model $\mathcal{M}$ is such that $v_{\mathcal{M}}$ satisfies the clause above, we will say that $\mathcal{M}$ is a *fixed-point for* $Par(x)$. It is important to note that not every model will be like this. In particular, there will be Kripke fixed-points (fixed-points for the truth predicate) in $\mathcal{M}^{ext}$ that are *not* fixed-points for $Par(x)$. These are needed to consistently interpret the paradoxicality predicate. For any model $\mathcal{M}$ that is a fixed-point for $Par(x)$, the following holds: $t^{\mathcal{M}} \in E_{\mathcal{P}}$ if and only if $t^{\mathcal{M}} = \ulcorner\phi\urcorner$ and $\forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2}$; and $t^{\mathcal{M}} \in A_{\mathcal{P}}$ if and only if: (i) $t^{\mathcal{M}} = \ulcorner\phi\urcorner$, and $\forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = 1$ or $\forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = 0$; or (ii) $t^{\mathcal{M}}$ is not the code of a statement.

In Rosenblatt (2021) it is shown with some detail how to set up a specific model satisfying these conditions, so here we will only offer a quick-and-dirty explanation of how the construction works. The general idea is to set up a sequence of models such that at each model one assigns a value to statements of the form $Par\ulcorner\phi\urcorner$, and then one runs Kripke's fixed-point construction for the truth predicate. The sequence starts with a model $\mathcal{M}_0$ which is the minimal fixed-point of the Kripke construction. In $v^{\mathcal{M}_0}$ every grounded statement is *either* true or false, and every statement of the form $Par\ulcorner\phi\urcorner$ is *neither*-true-nor-false. In other words, if $\mathcal{M}_0 = \langle\mathbb{N}, (E_{\mathcal{T}}, A_{\mathcal{T}}), (E_{\mathcal{P}}, A_{\mathcal{P}})\rangle$, then $E_{\mathcal{T}}$ is the set of statements that are true at Kripke's minimal fixed-point, $A_{\mathcal{T}}$ is the set of statements that are false at Kripke's minimal fixed-point, and $E_{\mathcal{P}} = A_{\mathcal{P}} = \varnothing$. After that one proceeds in stages. To calculate the truth-value of a statement of the form $Par\ulcorner\phi\urcorner$ at some successor stage $v^{\mathcal{M}_{\alpha+1}}$, one needs to determine the truth-value of the statement ϕ at every Kripke fixed-point extending the prior stage, $v^{\mathcal{M}_{\alpha}}$. After the truth-values of the statements of the form $Par\ulcorner\phi\urcorner$ are settled, one can carry out the usual Kripkean fixed-point construction to determine the truth-values of the rest of the statements of $\mathcal{L}_{+}$. When one is done with that, one moves to the next stage. This process is repeated at every successor stage. At limit stages one simply looks at the intersection of the previous models. At some stage a *minimal* fixed-point for $Par(x)$ is reached—a model $\mathcal{M}_{\alpha}$ such that $\mathcal{M}_{\alpha} = \mathcal{M}_{\alpha+1}$.[14] We call this model, $\mathcal{M}_{FP}$, and we call the set of models (Kripke fixed-points) extending it, $\mathcal{M}_{FP}^{ext}$. The model $\mathcal{M}_{FP}$ is such that for every statement ϕ, $v_{\mathcal{M}_{FP}}(Par\ulcorner\phi\urcorner) = 1$ if and only if $\forall\mathcal{M}' \in \mathcal{M}_{FP}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2}$.

The model $\mathcal{M}_{FP}$ and set $\mathcal{M}_{FP}^{ext}$ of Kripke fixed-points extending $\mathcal{M}_{FP}$ are of special interest for us. We submit that $\mathcal{M}_{FP}$ yields an appropriate interpretation for

[14] Cf. Rosenblatt (2021) for a more detailed presentation.

the paradoxicality predicate, and it does so by looking at the behaviour of statements at $\mathcal{M}_{FP}^{ext}$. In particular, we can identify the paradoxical statements as those that are in the extension of the paradoxicality predicate at $\mathcal{M}_{FP}$. In other words, a statement ϕ is *paradoxical* according to the present account if and only if $v_{\mathcal{M}_{FP}}(Par\ulcorner\phi\urcorner) = 1$; or, equivalently, if and only if $\forall\mathcal{M}' \in \mathcal{M}_{FP}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2}$. We can thus say, borrowing Kripke's words (Kripke 1975, p. 706), that what hitherto has been a meta-theoretically defined concept with no object-language counterpart, becomes a object-level predicate with specific semantic rules in the present theory.

3 Evaluating the fixed-point account

In this section, we will do two things. We will first consider how the account of paradoxicality that we have just described fares with respect to the desiderata discussed in Sect. 1. After that, we will highlight an aspect of the fixed-point conception that can be viewed as an important virtue, its flexibility.

3.1 General desiderata

To begin with, we submit that the account of paradoxicality that we have provided scores really well on the general desiderata discussed in Sect. 1. That the account has virtues such as consistency, naturalness, simplicity and explanatory power seems clear to us, so we will not spend too much space arguing for that here.

First, the construction offered in Sect. 2 reaches a fixed-point for the paradoxicality predicate (the model $\mathcal{M}_{FP}$) in roughly the same way that Kripke's construction reaches a fixed-point for the truth predicate. That fixed-point is such that no statement is both true and false at it, so the account is consistent.[15] *Second*, the characterization of $Par(x)$ is arguably a natural generalization of the definition of paradoxicality originally given by Kripke: a statement is paradoxical according to $\mathcal{M}_{FP}$ if it does not obtain a classical truth-value at every Kripke fixed-point extending $\mathcal{M}_{FP}$. Thus, it does not seem to be committed to any ad hoc hypothesis about the concept of paradoxicality. *Third*, the account is fairly simple too. It only employs a familiar possible world semantics wherein fixed-points play the role of possible worlds and the extension relation plays the role of the usual accessibility relation. As a consequence, many of the virtues of possible world semantics are preserved in this framework. *Fourth*, the account has explanatory power. One can extract from it an explanation

[15] Of course, the account will not be consistent if the underlying logic is paraconsistent. But paraconsistent logicians will suggest, first, that what is crucial is non-triviality rather than consistency, and, second, that consistency is just one among various other virtues that a theory may possess. Inconsistency can be viewed as a theoretical cost that can be trumped by other virtues. As Priest (2016, p. 351) puts it: "(...) it is only one criterion amongst many. How to weight it is, I am sure, itself the subject of some dispute. But whatever the weight, an inconsistent theory can be rationally preferable to a consistent one, if the performance of the inconsistent theory outweighs the consistent one on the other criteria."

of what makes a statement paradoxical (as we will discuss in more depth in the next subsection).[16]

3.2 Explaining what makes a statement paradoxical

As for the desiderata that are specific to theories of paradoxicality, we can start by noting that the account succeeds in offering an explanation of what makes a statement paradoxical. The paradoxicality of a statement ϕ can be identified with a specific model-theoretic fact, namely that ϕ does not obtain a classical truth-value at any Kripke fixed-point. So what makes a statement paradoxical is its behaviour across the set of Kripke fixed-points extending $\mathcal{M}_{FP}$.

Moreover, the theory we have offered is such that if ϕ is paradoxical in Kripke's original account, then $\ulcorner\phi\urcorner$ is in the extension of $Par(x)$ at $\mathcal{M}_{FP}$, i.e. $v_{\mathcal{M}_{FP}}(Par\ulcorner\phi\urcorner) = 1$. This does not mean that $Par(x)$ is extensionally equivalent to Kripke's original concept of paradoxicality. There are some statements ϕ containing occurrences of the paradoxicality predicate such that $Par\ulcorner\phi\urcorner$ is true at $v_{\mathcal{M}_{FP}}$, but those statements are not even expressible in Kripke's framework. For example, if ϕ is the statement $\neg Par\ulcorner 0 = 0\urcorner \wedge \lambda$, or the statement $\neg Par\ulcorner 0 = 0\urcorner \wedge (\lambda \vee Par\ulcorner 0 = 0\urcorner)$, $Par\ulcorner\phi\urcorner$ will be true at $v_{\mathcal{M}_{FP}}$. In this sense the present account goes beyond Kripke's.

Of course, we think that this is as it should be. Our goal is *not* to offer an account of paradoxicality that *extensionally* coincides with the account developed by Kripke. In particular, we are not requiring that something ought to be in the extension of $Par(x)$ at $\mathcal{M}_{FP}$ only if it is paradoxical according to Kripke's original account. For this, a classical and typed paradoxicality predicate would be more suitable. Our goal, rather, is to develop a theory which, as it were, *intensionally* coincides with Kripke's. That means that it ought to preserve the thought that a statement is paradoxical if and only if it does not obtain a classical truth-value at every relevant interpretation. Our account certainly achieves that by identifying the paradoxical statements as those that are in the extension of the predicate $Par(x)$ at $\mathcal{M}_{FP}$.

3.3 Principles for paradoxicality

We have suggested that the model $\mathcal{M}_{FP}$ yields an adequate characterization of the paradoxicality predicate. If one is interested in finding out what follows from what in virtue of this characterization, this model has to be associated with a notion of consequence. Following Rosenblatt (2021), one natural option in this setting is to equate consequence with truth-preservation at $\mathcal{M}_{FP}$.

Definition *($\mathcal{M}_{FP}$-Consequence)* A statement ϕ is an $\mathcal{M}_{FP}$-consequence of a set of statements Γ ($\Gamma \models \phi$) if, and only if, $v_{\mathcal{M}_{FP}}(\phi) = 1$ whenever $v_{\mathcal{M}_{FP}}(\gamma) = 1$ for every $\gamma \in \Gamma$. A statement ϕ is $\mathcal{M}_{FP}$-valid ($\models \phi$) if, and only if, $v_{\mathcal{M}_{FP}}(\phi) = 1$.

[16] There are other general theoretical virtues (see Footnote 6) that we are not taking into account. Some of them may not apply to theories of paradoxicality. We do not think this is a problem. It is not our intention to argue in favor of any form of anti-exceptionalism, so we are happy to admit that there may be some theoretical virtues that play an important role in the assessment of other types of scientific theories but that are of no significance for theories of paradoxicality.

It is not too hard to verify that an account based on this definition of consequence sanctions a number of principles connecting the notion of paradoxicality with other notions (cf. again Rosenblatt 2021). In particular, the paradoxicality predicate interacts with the logical connectives, the quantifiers and the truth predicate in the way one would expect. A statement is paradoxical if and only if its negation is paradoxical: $Par\ulcorner\phi\urcorner \models Par\ulcorner\neg\phi\urcorner$ and $Par\ulcorner\neg\phi\urcorner \models Par\ulcorner\phi\urcorner$. If two statements are paradoxical, so is their conjunction (though the converse fails): $Par\ulcorner\phi\urcorner \wedge Par\ulcorner\psi\urcorner \models Par\ulcorner\phi \wedge \psi\urcorner$, but $Par\ulcorner\phi \wedge \psi\urcorner \not\models Par\ulcorner\phi\urcorner \wedge Par\ulcorner\psi\urcorner$. If every instance of a universally quantified statement is paradoxical, so is the universal quantification (though the converse fails): $\forall x Par\ulcorner\phi(x)\urcorner \models Par\ulcorner\forall x\phi(x)\urcorner$ but $Par\ulcorner\forall x\phi(x)\urcorner \not\models \forall x Par\ulcorner\phi(x)\urcorner$. If a statement is true, it is not paradoxical (though the converse fails too): $Tr\ulcorner\phi\urcorner \models \neg Par\ulcorner\phi\urcorner$ but $\neg Par\ulcorner\phi\urcorner \not\models Tr\ulcorner\phi\urcorner$.[17]

Still, not everything is as one might have expected. Julien Murzi and Lorenzo Rossi have recently suggested that the notion of paradoxicality produces revenge paradoxes that affect paracomplete accounts of the truth-theoretic paradoxes.[18] Roughly, according to them, a case can be made that any paradoxicality predicate faithful to Kripke's theory should obey the following meta-rules[19]:

$$Par\text{-intro}\ \frac{\Gamma, \phi \vee \neg\phi \vDash \bot}{\Gamma \vDash Par\ulcorner\phi\urcorner} \qquad\qquad Par\text{-elim}\ \frac{\Gamma \vDash Par\ulcorner\phi\urcorner \qquad \Delta \vDash \phi \vee \neg\phi}{\Gamma, \Delta \vDash \bot}$$

The justification for Par-intro and Par-elim is that paracomplete theorists are arguably committed to the claim that a statement is paradoxical if and only if it satisfies excluded middle only on pain of triviality. The left-to-right direction of this claim justifies Par-elim, while the right-to-left direction justifies Par-intro. Murzi and Rossi show that, under fairly minimal assumptions, these rules lead to a contradiction.[20] To obtain a contradiction they rely on a self-referential statement, ρ, that says of itself that it is paradoxical if true, $Tr\ulcorner\rho\urcorner \rightarrow Par\ulcorner\rho\urcorner$.[21]

We need not look at the details of their argument, but we should briefly explain how the contradiction can be avoided in the present framework. As it is pointed out in Rosenblatt (2021), it is not difficult to check that, under the definition of consequence offered above, although the rule Par-elim holds in full generality, the rule Par-intro

---

[17] The verification of these claims is left to the reader. We employ the usual convention of writing for example $Par\ulcorner\neg\phi\urcorner$ instead of the more cumbersome $Par\ulcorner neg(\phi)\urcorner$ (where neg represents the corresponding function operating on codes of statements). Also, we write $\forall x Par\ulcorner\phi(x)\urcorner$ instead of $\forall x Par\ulcorner\phi\urcorner(\dot{x}/\ulcorner x\urcorner)$ (where $\dot{x}$ represents the function that maps each number to its numeral and $Par\ulcorner\phi\urcorner(\dot{x}/\ulcorner x\urcorner)$ is the claim that the result of substituting the numeral of x for the variable x in the statement ϕ is paradoxical).

[18] Actually, they are in fact targeting not only paracomplete approaches but also paraconsistent, non-transitive, non-contractive and non-reflexive approaches. Their arguments are developed in Murzi and Rossi (2020) and Murzi and Rossi (2022).

[19] Disjunction, $\vee$, can be defined using conjunction and negation in the usual way: $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$. We assume that the falsity constant, $\bot$, is such that for every model $\mathcal{M}$, $v_{\mathcal{M}}(\bot) = 0$.

[20] The paradox also requires a number of logical rules and also 'recapture' rules, that is, rules establishing that one can reason classically in certain contexts. But these are rules that paracomplete theorists typically accept. (With one exception: the argument requires the rule of disjunction-introduction, so it does not obviously carry over to a paracomplete theory based on the weak Kleene schema.)

[21] The conditional, $\rightarrow$, can be defined in the following way: $\phi \rightarrow \psi := \neg(\phi \wedge \neg\psi)$. It would be interesting to see how this paradox plays out with an intensional conditional, of the sort that Field, Priest and others have advocated, but we leave a careful study of this possibility for another occasion.

fails. There are statements ϕ such that $\phi \vee \neg\phi$ entails a contradiction, but $Par\ulcorner\phi\urcorner$ does not hold. The Murzi-Rossi statement, ρ, behaves exactly in that way. In order to evaluate $Par\ulcorner\rho\urcorner$ at $v_{\mathcal{M}_{FP}}$ we need to consider the behaviour of ρ across all the Kripke fixed-points $\mathcal{M}$ extending $\mathcal{M}_{FP}$. Clearly, in virtue of the definition of $\rightarrow$, for any such model $\mathcal{M}$, it holds that $v_{\mathcal{M}}(\rho) = 1$ just in case $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$, and $v_{\mathcal{M}}(\rho) = \frac{1}{2}$ just in case $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) \in \{0, \frac{1}{2}\}$. It follows that there is no model $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(\rho) = 0$, but there are models $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(\rho) = \frac{1}{2}$, and models $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(\rho) = 1$. From this, we can infer the following two facts: (i) that $v_{\mathcal{M}_{FP}}(\rho) = \frac{1}{2}$, and thus, by the definition of $\vee$, $v_{\mathcal{M}_{FP}}(\rho \vee \neg\rho) = \frac{1}{2}$; and (ii) that, by the semantic clause for $Par(x)$, $v_{\mathcal{M}_{FP}}(Par\ulcorner\rho\urcorner) = \frac{1}{2}$. Applying the definition of consequence, we conclude that $\rho \vee \neg\rho \models \bot$ and $\not\models Par\ulcorner\rho\urcorner$, so Par-intro fails for ρ.

Crucially, some of the models that play a role here are *not* fixed-points for $Par(x)$ and thus should not be understood as yielding an adequate account of the extension of the paradoxicality predicate. They are nonetheless necessary to determine the truth-value of $Par\ulcorner\rho\urcorner$ at $v_{\mathcal{M}_{FP}}$. In particular, there are models $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(\rho) = v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$. These are models where ρ is both true and paradoxical! Even though $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$, it is not the case that for every $\mathcal{M}'$ such that $\mathcal{M} \preccurlyeq \mathcal{M}'$, $v_{\mathcal{M}'}(\rho) = \frac{1}{2}$. In fact, since for every model $\mathcal{M}$, $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$ just in case $v_{\mathcal{M}}(\rho) = 1$, we have that for every $\mathcal{M}'$ such that $\mathcal{M} \preccurlyeq \mathcal{M}'$, $v_{\mathcal{M}'}(\rho) = 1$. There are also models $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 0$ and models $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = \frac{1}{2}$. Some of these models will be maximal, in the sense that the only model extending $\mathcal{M}$ will be $\mathcal{M}$ itself. Given that for every model $\mathcal{M}$, $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) \in \{0, \frac{1}{2}\}$ just in case $v_{\mathcal{M}}(\rho) = \frac{1}{2}$, these models will not be fixed-points for $Par(x)$ either, since, trivially, for every $\mathcal{M}'$ such that $\mathcal{M} \preccurlyeq \mathcal{M}'$, $v_{\mathcal{M}'}(\rho) = \frac{1}{2}$, but $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) \neq 1$.

In spite of this, there is an interesting class of instances of Par-intro that are warranted. Par-intro holds for a statement ϕ if ϕ is grounded or paradoxical in Kripke's sense. In particular, for every liar-like statement ϕ, one has $\models Par\ulcorner\phi\urcorner$. On account of this, the unavailability of Par-intro for some purportedly revenge-generating statements need not be seen as a serious drawback.

However, one could suggest, first, that the notion of consequence employed in Rosenblatt (2021) is not sufficiently general in that it relies on a single model; and second, that to restrict Par-in amounts to forsake the naive notion of paradoxicality, a cost that the paracomplete theorist should not be willing to pay. If so, a more general and naive-friendly definition of consequence ought to be used. One natural thought—not considered in Rosenblatt (2021)—is to quantify over the set $\mathcal{M}_{FP}^{ext}$ of Kripke fixed-points extending $\mathcal{M}_{FP}$.[22] More formally:

[22] There is a different way in which the definition can be generalized. One could consider models $\mathcal{M}$ for the base language $\mathcal{L}_{PA}$ other than the standard model of PA, and then quantify over every fixed-point extending each of the minimal fixed-points for $Par(x)$ that can be reached from each of those 'base' models. This is useful if, for example, one wishes to give a diagnosis of contingent paradoxes. However, since we are limiting ourselves to the standard model of PA, we will not consider this possibility here. Needless to say, nothing important hangs on this (cf. Gallovich and Rosenblatt 2022 for the details).

Definition *($\mathcal{M}^{ext}_{FP}$-Consequence)* A statement ϕ is an $\mathcal{M}^{ext}_{FP}$-consequence of a set of statements Γ ($\Gamma \models^* \phi$) if, and only if, for every $\mathcal{M} \in \mathcal{M}^{ext}_{FP}$, $v_{\mathcal{M}}(\phi) = 1$ whenever $v_{\mathcal{M}}(\gamma) = 1$ for every $\gamma \in \Gamma$. A statement ϕ is $\mathcal{M}^{ext}_{FP}$-valid ($\models^* \phi$) if, and only if, for every $\mathcal{M} \in \mathcal{M}^{ext}_{FP}$, $v_{\mathcal{M}}(\phi) = 1$.

The difference between $\mathcal{M}_{FP}$-consequence and $\mathcal{M}^{ext}_{FP}$-consequence is that the former only considers $\mathcal{M}_{FP}$, whereas the latter quantifies over every fixed-point extending $\mathcal{M}_{FP}$. In a way, the distinction between $\models^*$ and $\models$ is reminiscent of the distinction between local consequence (truth-preservation) and global consequence (validity-preservation), which is typical in other logical frameworks, such as supervaluationism and modal logic. On the one hand, $\models^*$ is a local notion because it requires, for every model, truth-preservation at that model. On the other hand, $\models$ is a global notion because for any statement ψ, $v_{\mathcal{M}_{FP}}(\psi) = 1$ if, and only if, for every $\mathcal{M} \in \mathcal{M}^{ext}_{FP}$, $v_{\mathcal{M}}(\psi) = 1$. This means that there is an equivalent way of stating the definition of $\mathcal{M}_{FP}$-consequence: a statement ϕ is an $\mathcal{M}_{FP}$-consequence of a set of statements Γ ($\Gamma \models \phi$) if, and only if, for every $\mathcal{M} \in \mathcal{M}^{ext}_{FP}$, $v_{\mathcal{M}}(\phi) = 1$ whenever for every $\mathcal{M} \in \mathcal{M}^{ext}_{FP}$, $v_{\mathcal{M}}(\gamma) = 1$ for every $\gamma \in \Gamma$. Simply put, $\models$ is a global notion because it requires validity-preservation.

If one endorses $\mathcal{M}^{ext}_{FP}$-Consequence, then Par-intro holds unrestrictedly, but Par-elim fails. That is, there are cases where $\Gamma \models^* \phi \vee \neg\phi$ and $\Delta \models^* Par\ulcorner\phi\urcorner$, but $\Gamma, \Delta \not\models^* \bot$. The Murzi-Rossi statement ρ behaves exactly in that way. As we have seen, there is a (Kripke) fixed-point $\mathcal{M}$ extending $\mathcal{M}_{FP}$ such that $v_{\mathcal{M}}(\rho) = v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$. It follows that $\rho \vee \neg\rho \not\models^* \bot$. However, since $v_{\mathcal{M}}(\rho) = 1$ if and only if $v_{\mathcal{M}}(Par\ulcorner\rho\urcorner) = 1$, and for no $\mathcal{M}$, $v_{\mathcal{M}}(\rho) = 0$, we have $\rho \vee \neg\rho \models^* Par\ulcorner\rho\urcorner$. Given that $\rho \vee \neg\rho \models^* \rho \vee \neg\rho$ is also the case, we have a counterexample to Par-elim.

Once again, we think that this need not be seen as a serious drawback. It is only some instances of Par-elim that fail, and in fact in this case one can even retain the following version of Par-elim:

$$Par\text{-elim}^* \ \frac{\models^* Par\ulcorner\phi\urcorner \qquad \Delta \models^* \phi \vee \neg\phi}{\Delta \models^* \bot}$$

On the one hand, if $\models^* Par\ulcorner\phi\urcorner$, then for every Kripke fixed-point $\mathcal{M}$ extending $\mathcal{M}_{FP}$, $v_{\mathcal{M}}(\phi) = \frac{1}{2}$. On the other hand, if $\Delta \models^* \phi \vee \neg\phi$, then for every Kripke fixed-point $\mathcal{M}$ extending $\mathcal{M}_{FP}$, if $v_{\mathcal{M}}(\delta) = 1$ for every $\delta \in \Delta$, then $v_{\mathcal{M}}(\phi) \in \{1, 0\}$. Thus, the premises of Par-elim cannot be jointly satisfied under the assumption that the statements in Δ are true, which means that Par-elim* holds.

The upshot is that one can have natural-looking introduction and elimination rules for $Par(x)$ if one employs the notion of $\mathcal{M}^{ext}_{FP}$-consequence. The rule Par-elim* is sufficient to yield (a restricted version of) the bottom-to-top direction of Par-intro: from $\models^* Par\ulcorner\phi\urcorner$ one can infer $\phi \vee \neg\phi \models^* \bot$, using $\phi \vee \neg\phi \models^* \phi \vee \neg\phi$ as a side premise. This means that one can retain the paracomplete theorist's naive idea that a statement is paradoxical if and only if it satisfies excluded middle only on pain of triviality. That is, $\models^* Par\ulcorner\phi\urcorner$ if and only if $\phi \vee \neg\phi \models^* \bot$.[23]

[23] One potential cost of relying on the notion of $\mathcal{M}^{ext}_{FP}$-consequence is that one looses some of the facts alluded to earlier pertaining to the interaction of the paradoxicality predicate with the logical connectives,

We will not try to defend one of these notions of consequence over the other. We want to remain as non-committal as possible. In particular, $\models^*$ should not be thought of as an improvement over $\models$, but rather as an additional option available to the fixed-point theorist. Depending on which of Par-intro and Par-elim one thinks is less costly to restrict, one will favor one notion or the other. All we mean to suggest is that the paracomplete theorist has tools at her disposal to avoid the revenge argument involving ρ while retaining a strong notion of paradoxicality.

3.4 A unified account of truth and paradox

Another reasonable desideratum for a theory of truth and paradoxicality is that the truth-theoretic paradoxes and the paradoxes involving the notion of paradoxicality must receive a similar treatment. For example, the statements λ and ρ are constructed using exactly the same diagonal technique, and they appear to be structurally similar, so it is natural to expect a unified treatment. In our approach, however, one can assert that λ is paradoxical, but not that ρ is paradoxical. The worry, then, is that the approach fails to explain in what way λ and ρ have the same defective semantic status.

Alas, things are not so simple. We think that there are various ways of understanding the idea of a "unified treatment" of the paradoxes. To unpack it, let's view it from three different perspectives. It will turn out that under each of these perspectives, there is sense in which our approach does in fact offer a unified treatment of the paradoxes.[24]

If one puts a lot of weight on the idea that truth and paradoxicality ought to be interpreted along similar (naive) lines—in that one should seek to retain naive semantic concepts in the face of the paradoxes—then restricting Par-elim in the manner suggested in the previous section seems to be the way to go. The point is that the restriction on Par-elim is compatible with naivety for Par, understood as the claim that for any statement ϕ, $\phi \vee \neg\phi \models^* \bot$ if and only if $\models^* Par\ulcorner\phi\urcorner$. In particular, it is possible to prove that the revenge statement ρ satisfies the relevant biconditional, since $\rho \vee \neg\rho \not\models^* \bot$ and $\not\models^* Par\ulcorner\rho\urcorner$. Thus, one can hold on to the idea that λ and ρ receive the same treatment. In both cases one retains naivety, since Par-intro and Par-elim* are already sufficient to yield a naive paradoxicality predicate.

If, instead, one thinks that a restriction on Par-intro is more in keeping with the paracomplete theorist's overall account, then one should give up on the notion of naive paradoxicality. The rejection of Par-intro for ρ—it will be the case that $\rho \vee \neg\rho \models \bot$ and $\not\models Par\ulcorner\rho\urcorner$—harmonizes well with the paracomplete theorist's rejection of Reductio for λ and her rejection of Conditional Proof for the Curry sentence. The paradoxicality of a statement ϕ can still be identified with the model-theoretic idea that ϕ is $\frac{1}{2}$ at every Kripke fixed-point, but this idea does not coincide with $\phi \vee \neg\phi$ entailing a contradiction. Yet, to the extent that the failure of Par-intro is nicely in

Footnote 23 continued
the quantifiers and the truth predicate. For example, it is easy to see that ρ yields a counterexample to $Tr\ulcorner\phi\urcorner \models^* \neg Par\ulcorner\phi\urcorner$.

24 We are indebted to Luca Incurvati, Julien Murzi, Lorenzo Rossi and Giorgio Sbardolini for discussion on these ideas.

keeping with the paracomplete theorist's general take on introduction rules, she can still claim that she is offering a unified approach to λ and ρ.[25]

An altogether different possibility is to deny that λ and ρ give rise to paradoxes of the same kind. As we pointed out, in our account one can assert that λ is paradoxical, but not that ρ is paradoxical. So there is a clear sense in which λ and ρ are diagnosed differently. The idea is that our characterization of the paradoxicality predicate in terms of Kripke fixed-points implies the existence of models where ρ is true, so our way of understanding $Par(x)$ reveals that the seeming paradoxicality of ρ vanishes once one interprets this statement properly. Strictly speaking, ρ was not a paradox after all, and our account adequately reflects this fact. Therefore, the requirement that one ought to offer a unified treatment of ρ and λ does not apply, and the paracomplete theorist is justified in treating them differently.[26]

We have offered three different perspectives on how to understand the idea of a unified treatment of the paradoxes. Once again, for the purposes of the paper we want to remain as non-committal as possible. We think that each of these options has advantages, as well as costs. The crucial point is that regardless of the option one ultimately endorses, there is a sense in which it is reasonable to say that our account abides by the requirement that the paradoxes involving the paradoxicality predicate must be treated in the same way that the truth-theoretic paradoxes.[27]

3.5 The fixed-point conception and its flexibility

Before moving on to consider one limitation that affects the theory we have presented, it is important to mention that our proposal is just one possible exemplar of the fixed-point conception. By this we mean that there are many ways of capturing the central thought underpinning the fixed-point conception and our account embodies just one of them. We believe that this is one of the most appealing features of this conception, its flexibility. Given that the characterization of the paradoxicality predicate only relies on

25 This is roughly the way in which Rosenblatt (2021) justifies the imposition of a restriction on Par-intro.

26 If ρ is not really a paradox, then one may ask if there are any new paradoxical statements involving the notion of paradoxicality. As one can infer from Sect. 3.2, the answer to this question is positive, although these new paradoxes are not very interesting. For example, the statement $\neg Par\ulcorner 0 = 0 \urcorner \wedge \lambda$ is a paradox of this kind and it is diagnosed as such by our account.

27 At this point, it is important to stress that it is not our aim to establish that the approach we are offering is revenge-free in general. Considering the discussion given by Murzi and Rossi (2020), there is an important distinction one can draw between object-linguistic revenge paradoxes and meta-theoretic revenge paradoxes. They say that object-linguistic revenge paradoxes point to the inexpressibility in a theory of some notion that plays an explanatory or expressive role in that theory, while meta-theoretic revenge paradoxes involve notions that can be defined in the (classical) meta-theory. Given that the idea of 'playing an explanatory or expressive role' is one that does not admit of a formal characterization, we think that revenge-freedom is not something that can be formally or conclusively established. Whether a theory is revenge-free in the relevant sense will crucially depend on whether the notions that are inexpressible in the theory are notions that play an explanatory or expressive role in it. In the case of paradoxicality, it seems hard to deny that the notion plays an explanatory role in various non-classical theories. What we have shown is that one can consistently represent that notion in the object-language of a paracomplete theory as long as one is willing to slightly weaken one of the rules Par-intro and Par-elim. To be sure, that only yields revenge-freedom if (i) the resulting paradoxicality predicate is sufficient to play the explanatory role the non-classical theorist expects it to play, and (ii) there are no other inexpressible notions that play an explanatory or expressive role in the non-classical theorist's overall picture.

fixed-point models, the conception is compatible with several theories that are based on these models and thus with various different solutions to the semantic paradoxes.

For one thing, one may change the underlying evaluation schema. We have defined the logical expressions using strong Kleene interpretations, but there are no technical difficulties with using weak Kleene interpretations or various types of supervaluations. The characterization of the paradoxicality predicate is compatible with these alternatives. For another, we have interpreted the value $\frac{1}{2}$ as neither-true-nor-false, so one natural thought is to define the notion of validity in terms of preservation of value 1 (i.e., preservation of truth). But there are various other options available. For example, one could interpret the value $\frac{1}{2}$ as both-true-and-false and define validity in terms of preservation of both 1 and $\frac{1}{2}$. This would yield a paraconsistent theory of paradoxicality.[28] It is also possible to modify the definition of validity to obtain yet other types of solutions, such as those based on non-transitive or non-reflexive consequence relations. What makes all these variations possible is that the characterization of paradoxicality is invariant under permutations of the definition of validity. Regardless of how validity is defined, the paradoxical statements are those that lack a classical value at every fixed-point.

Still, it may be suggested that all these approaches are non-classical. This could be a problem: if the fixed-point approach rules out classical logic, it is not as flexible as it should be. However, we submit that to the extent that Kripke's original account can be interpreted in a way that makes it compatible with classical logic, this approach can too. Kripke himself thinks that there is nothing intrinsically non-classical about his theory of truth.[29] Although, in our opinion, time has shown that the non-classical interpretation of Kripke's theory is more interesting and fruitful, that does not mean that the classical interpretation is incoherent. In fact, the usefulness of fixed-point models is not decreased if one thinks that they are mere conventions for handling sentences that fail to express a proposition. A similar thought, we believe, applies to our account of paradoxicality. One can use fixed-point models to characterize paradoxicality even if paradoxical sentences are among the sentences that fail to express a proposition. In view of this, it is fair to say that the classical logician can employ the apparatus of fixed-points to offer a characterization of paradoxicality in much the same way that the non-classical logician can.[30]

[28] One should be careful, though. Since $\frac{1}{2}$ is now designated, the theory will be such that $\models Par\ulcorner\phi\urcorner$ in some cases where ϕ is not paradoxical in Kripke's account. At any rate, it is still true that (i) ϕ is paradoxical in Kripke's account only if $Par\ulcorner\phi\urcorner$ is *strictly* true and (ii) ϕ is not paradoxical in Kripke's account only if $\models \neg Par\ulcorner\phi\urcorner$.

[29] As he puts it: "conventions for handling sentences that do not express propositions are not in any philosophically significant sense "changes in logic." The term 'three-valued logic', occasionally used here, should not mislead." Cf. Kripke (1975, fn. 18).

[30] There may be another sense in which the fixed-point account is compatible with classical logic. It is well known that one can obtain a classical theory of truth from Kripke models if one takes the 'close-off' of these models. That is, one takes whatever statements lie outside the extension of the truth predicate at some fixed-point and then one stipulates that $Tr(x)$ is false of those statements. Kripke himself notes this possibility in the *Outline*. It would be interesting to explore if the same can be done with the construction we have offered for $Par(x)$.

4 Limitations

We have not yet discussed one of the desiderata that we introduced for a theory of paradoxicality—namely, that the theory should agree with our intuitions in a large number of cases. The matter is somewhat complicated because it is controversial that we even have clear intuitions about paradoxicality, but let us assume that we do. If so, then it is uncontentious to say that the theory we have offered agrees with our intuitions in a large number of cases. It not only establishes the paradoxicality of the liar sentence, but the paradoxicality predicate applies exactly to the statements that behave in a liar-like way.

However, even if the theory does a fairly good job at capturing various intuitions, it faces two substantial objections. On the one hand, there are paradoxical statements (in Kripke's sense) that arguably should not be identified as paradoxical—this is an old objection to Kripke's theory due to Gupta. On the other hand, there are statements that fail to be paradoxical (in Kripke's sense) for which the theory remains silent, in that it neither says that they are paradoxical nor that they are not. In this section we will analyze both objections.

Let's first consider Gupta's objection.[31] Gupta argues that Kripke's definition of the concept of paradox is counterintuitive because it entails that some logical laws (of classical logic) are paradoxical. Of course, our Kripke-inspired theory has to face this objection as well. Consider a claim saying that no statement is both true and untrue, $\forall x \neg(Tr(x) \wedge \neg Tr(x))$. According to Kripke's theory based on the strong or weak Kleene schemata, this statement is paradoxical—there is no fixed-point at which it obtains a classical truth-value. So it follows that $Par\ulcorner\forall x \neg(Tr(x) \wedge \neg Tr(x))\urcorner$ is true at $\mathcal{M}_{FP}$. Is this diagnosis misguided?

Here are two tentative thoughts. First, the example only applies to certain evaluation schemata. For instance, if one endorses a supervaluational schema, then the statement is not paradoxical, because it is an instance of a validity of first-order classical logic. Secondly, if one goes for one of Kleene's schemata, we think that it should not be too surprising that $\forall x \neg(Tr(x) \wedge \neg Tr(x))$ comes out as paradoxical, since this statement can be understood as an infinite conjunction wherein one of the conjuncts, $\neg(Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner)$, is equivalent to λ. What this shows, then, is something that we sort of already knew, namely that some statements that are valid in classical logic turn out to be paradoxical in the non-classical setting. Of course, this does not mean that there is no clash with natural language, there is one for sure. But we think that this is probably due to the fact that our natural language-based intuitions are fueled only by the non-pathological instances of $\forall x \neg(Tr(x) \wedge \neg Tr(x))$. Once one realizes that 'x' can be instantiated with pathological statements, the intuition of the non-paradoxicality of $\forall x \neg(Tr(x) \wedge \neg Tr(x))$ loses some of its force.

In our framework (based on the strong Kleene schema) these facts can be captured in a very straightforward way. On the one hand, there are ϕs such that we

[31] Gupta (1982) offers a number of other objections to the fixed-point approach. Since they are not specifically related to paradoxicality, we have decided to omit them. But, to be sure, a full defense of the fixed-point approach to truth and paradox would require an answer to these other points as well.

have $\exists x Par\ulcorner\phi(x)\urcorner \not\models Par\ulcorner\forall x\phi(x)\urcorner$.[32] But, if $\phi(x)$ is $\neg(Tr(x) \wedge \neg Tr(x))$ or any other validity of classical logic, then $\exists x Par\ulcorner\phi(x)\urcorner \models Par\ulcorner\forall x\phi(x)\urcorner$, because the paradoxicality of one instance is sufficient for the paradoxicality of the universally quantified statement. Also, for the same reason, one can assert $\forall x\neg(Tr(x) \wedge \neg Tr(x))$ under the assumption that no instance of it is paradoxical. In other words, we have: $\forall x\neg Par\ulcorner\neg(Tr(x) \wedge \neg Tr(x))\urcorner \models \forall x\neg(Tr(x) \wedge \neg Tr(x))$.

In order to analyze the second objection in detail, we will take a bit of a detour. It will be useful to introduce a few semantic categories that can be extracted from the apparatus of Kripke fixed-points. These categories will allow us to separate the statements of the object-language into several different sets. First, statements are either grounded or ungrounded. Assuming, to simplify matters, that the statements of the base language (i.e., the Tr-free statements) always obtain a classical truth-value (either the value 1 or the value 0) at every fixed-point, one can define a *grounded* statement as a statement that is either true at the minimal fixed-point of Kripke's construction or false at the minimal fixed-point of Kripke's construction. For example, the statements $3 + 7 = 10$ and $Tr\ulcorner 2 + 9 = 10\urcorner$ will both be grounded. The class of ungrounded statements is much more diverse, and there are at least two different and non-equivalent ways of classifying them.

Simple classification Apart from *paradoxical* statements, i.e. statements that are $\frac{1}{2}$ at every fixed-point, there are statements that are ungrounded but unparadoxical. We will say that a statement is *hypodoxical* at a fixed-point if it is true at some fixed-points that extend it and false at some fixed-points that extend it.[33] Truth-theorists often consider a dual of λ, called the truth-teller. This is a statement, τ, saying of itself that it is true, $Tr\ulcorner\tau\urcorner$. τ is not paradoxical—at least not in the way that λ is—but it is hypodoxical at certain fixed-points. For example, since there are non-minimal fixed-points that make it true and other non-minimal fixed-points that make it false, τ is hypodoxical at the minimal fixed-point of Kripke's construction. There are yet other important classes of statements that can be characterized on the basis of their semantic behavior across fixed-points. Let's say that a statement is *s-true* (or *sometimes true*) at a fixed-point if there is a fixed-point that extends it at which it is true and there is also a fixed-point that extends it at which it is $\frac{1}{2}$, but there is no fixed-point that extends it at which it is false. Analogously, we will say that a statement is *s-false* at a fixed-point if there is a fixed-point that extends it at which it is false and there is also a fixed-point that extends it at which it is $\frac{1}{2}$, but there is no fixed-point that extends it at which it is true. To illustrate, if one disjoins the truth-teller with its negation, one obtains the statement $\tau \vee \neg\tau$. At the minimal fixed-point of Kripke's contruction, $\tau \vee \neg\tau$ is s-true, since there are non-minimal fixed-points where it is true. If one instead conjoins the truth-teller with its negation, one obtains the statement $\tau \wedge \neg\tau$. At the minimal fixed-point of Kripke's construction, $\tau \wedge \neg\tau$ is s-false, since there are non-minimal fixed-points where it is false.

The idea underpinning this classification is that the status of a statement at a fixed-point depends on its behaviour across the set of all fixed-points extending it. Following

[32] The existential quantifier, $\exists$, is not part of the official language, but it can be defined in the following way: $\exists x\phi(x) := \neg\forall x\neg\phi(x)$.

[33] In speaking of 'hypodoxes' we are following Peter Eldrige-Smith's terminology (Eldridge-Smith 2007).

this idea, there are four categories to which ungrounded statements of the language can belong: paradoxical, hypodoxical, s-true and s-false. These categories are jointly exhaustive—i.e., they jointly exhaust the set of ungrounded statements—and mutually exclusive—i.e., if a statement belongs to one of these categories it cannot belongs to any other category.

In a number of recent papers, Roy Cook and Nicholas Tourville have offered a different way to classify statements on the basis of their behaviour across Kripke fixed-points.[34]

Sophisticated classification Paradoxical statements are treated as in **Simple Classification**. But for other statements there are a number of subtle differences. A statement is said to be *semi-true* at a fixed-point $\mathcal{M}$ if for every fixed-point $\mathcal{M}'$ extending it there is a fixed-point $\mathcal{M}''$ extending $\mathcal{M}'$ that makes it true. A statement is *semi-false* at a fixed-point $\mathcal{M}$ if for every fixed-point $\mathcal{M}'$ extending it there is a fixed-point $\mathcal{M}''$ extending $\mathcal{M}'$ that makes it false. Also, a statement is *semi-classical* at a fixed-point $\mathcal{M}$ if for every fixed-point $\mathcal{M}'$ extending it, there is a fixed point $\mathcal{M}''$ extending $\mathcal{M}'$ that makes the statement either true or false. Finally, a statement is *unstable* at a fixed-point if it is ungrounded but it is neither of the above.

The idea underpinning this classification is that each of the categories involved gives rise to a monotone operator (in a novel technical sense defined by Cook and Tourville)[35]. According to this, there are five categories to which ungrounded statements of the language can belong: paradoxical, semi-true, semi-false, semi-classical and unstable. The resulting categories are jointly exhaustive for ungrounded statements, but not mutually exclusive—e.g., if a statement is semi-true or semi-false, it is also semi-classical.

Let us see how the differences between the two classifications play out in specific cases. First, the notions of hypodoxicality and semi-classicality are not equivalent. Focusing on the minimal fixed-point of Kripke's construction, there are statements that are semi-classical but not hypodoxical at that fixed-point, such as $\tau \vee \neg\tau$ and $\tau \wedge \neg\tau$ (and even grounded statements like $0 = 0$ and $0 = s(0)$, since a statement is semi-classical if it is grounded). Also, there are statements that are hypodoxical, but not semi-classical, such as $\tau_1 \vee (\lambda \wedge \tau_2)$ (where τ_1 and τ_2 are two independent truth-tellers). The reason is that there are maximal fixed-points where $\tau_1 \vee (\lambda \wedge \tau_2)$ is $\frac{1}{2}$—namely, those where τ_1 is false and τ_2 is true. So it is not the case that for every fixed-point there is another fixed-point extending it where this statement obtains a classical truth-value. Second, the notions of s-truth and semi-truth are not equivalent, either. There are statements that are s-true at the minimal fixed-point of Kripke's construction but not semi-true, such as the statement $\lambda \vee \tau$. There are maximal fixed-points where $\lambda \vee \tau$ is $\frac{1}{2}$—namely, those where τ is false. So it is not the case that for every fixed-point there is a fixed-point extending it where this statement is true. Also, there are statements that are semi-true at the minimal fixed-point of Kripke's construction but not s-true, such as $0 = 0$. The reason is that whereas it is the case that for every fixed-point there is a fixed-point extending it at which this statement is true

34 Cf. Cook and Tourville (2016), Cook and Tourville (2020) and, especially, Cook (2020).

35 Cook and Tourville's idea is that an operator or a predicate is monotonic if and only if it is intensionally monotonic, i.e. monotone relative to the order of the different intensional semantic statuses that a statement can have (which is an order defined by them). Cf. Cook (2020) for the details.

(since it is true at every fixed-point), there is no fixed-point at which this statement is $\frac{1}{2}$. For similar reasons, s-falsity and semi-falsity do not coincide. Finally, note that the statements $\tau_1 \vee (\lambda \wedge \tau_2)$ and $\lambda \vee \tau$ are both unstable at the minimal fixed-point of Kripke's construction according to **Sophisticated Classification**. Yet, according to **Simple Classification**, $\tau_1 \vee (\lambda \wedge \tau_2)$ is hypodoxical and $\lambda \vee \tau$ is s-true.

In a sense, neither classification is ideal. On the one hand, **Simple Classification** lumps together hypodoxical statements that are either true or false at every maximal fixed point, such as τ, and hypodoxical statements that are also $\frac{1}{2}$ at some of them, such as $\tau_1 \vee (\lambda \wedge \tau_2)$. **Simple Classification** also fails to distinguish between s-true statements that are true at every maximal fixed point, such as $\tau \vee \neg\tau$, and s-true statements that are $\frac{1}{2}$ at some of them, such as $\lambda \vee \tau$. Similarly, the classification fails to separate between s-false statements that are false at every maximal fixed point, such as $\tau \wedge \neg\tau$, and s-false statements that are $\frac{1}{2}$ at some maximal fixed points, such as $\lambda \wedge \tau$. On the other hand, **Sophisticated Classification** fails to make a distinction between unstable statements that behave very differently. Consider again the statements $\tau_1 \vee (\lambda \wedge \tau_2)$, $\lambda \vee \tau$, and $\lambda \wedge \tau$. There are maximal fixed points at which $\tau_1 \vee (\lambda \wedge \tau_2)$ is true, there are maximal fixed points at which it is false, and there are maximal fixed points at which it is $\frac{1}{2}$. However, there is no maximal fixed point at which $\lambda \vee \tau$ is false and there is no maximal fixed point at which $\lambda \wedge \tau$ is true. Despite these differences, **Sophisticated Classification** lumps the three statements together by deeming them all unstable.

With these differences in mind, we can now go back to our characterization of the paradoxicality predicate. It is a fact about our account that $\vDash Par\ulcorner\lambda\urcorner$, but $\nvDash Par\ulcorner\phi\urcorner$ for every unparadoxical statement. So there is a sense in which the paradoxicality predicate allows us to distinguish the liar from other statements without bringing meta-linguistic resources into play. However, one obvious limitation with the account is that it is not possible to express that hypodoxical statements are not paradoxical in the object language. In fact, for every hypodoxical statement ϕ, $\neg Par\ulcorner\phi\urcorner$ has the value $\frac{1}{2}$ at $v_{\mathcal{M}_{FP}}$, which means that the claim that these statements are unparadoxical does not hold. By way of example, the truth-teller, τ, behaves exactly in this way. It is not the case that τ is $\frac{1}{2}$ at every interpretation extending $\mathcal{M}_{FP}$, but $v_{\mathcal{M}_{FP}}(Par\ulcorner\tau\urcorner) = \frac{1}{2}$, so $\nvDash \neg Par\ulcorner\tau\urcorner$. Thus, even though the characterization gets the extension of $Par(x)$ right, some statements that fail to be paradoxical (in Kripke's sense) are not a part of its anti-extension.

One can try to amend this situation by changing the falsity conditions of statements of the form $Par\ulcorner\phi\urcorner$.[36] Then:

$$v_{\mathcal{M}}(Par\ulcorner\phi\urcorner) = \begin{cases} 1 & \text{if } \forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2} \\ 0 & \text{if } \forall\mathcal{M}' \in \mathcal{M}^{ext} : \exists\mathcal{M}'' \in \mathcal{M}^{ext} : \mathcal{M}' \preccurlyeq \mathcal{M}'' \\ & \quad \text{and } v_{\mathcal{M}''}(\phi) = 1 \text{ or } v_{\mathcal{M}''}(\phi) = 0 \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

[36] The characterization we are about to offer is due to Cook and Tourville. Since **Sophisticated Classification** differs from **Simple Classification** in its treatment of unparadoxical statements, it can be used to offer a different specification of the falsity conditions for paradoxicality claims.

According to this definition, $Par\ulcorner\phi\urcorner$ is false at $v_{\mathcal{M}}$ if every fixed-point in $\mathcal{M}^{ext}$ can be extended to one where ϕ is true or false. Equivalently, $Par\ulcorner\phi\urcorner$ is false at $v_{\mathcal{M}}$ if ϕ is semi-classical at $v_{\mathcal{M}}$. This (partially) solves the problem we posed before.

To see this let $\mathcal{M}^{\star}_{FP}$ be the model that is the *minimal* fixed-point for $Par(x)$ employing this alternative definition, and let $\mathcal{M}^{\star ext}_{FP}$ be the set of Kripke fixed-points extending $\mathcal{M}^{\perp}{}_{FP}$. Thanks to Zorn's Lemma, we know that every fixed-point can be extended to a maximal fixed-point. Given that τ is either true or false at every maximal fixed-point, it follows that $\forall \mathcal{M}' \in \mathcal{M}^{\star ext}_{FP} : \exists \mathcal{M}'' \in \mathcal{M}^{\star ext}_{FP} : \mathcal{M}' \preccurlyeq \mathcal{M}''$ and $v_{\mathcal{M}''}(\tau) = 1$ or $v_{\mathcal{M}''}(\tau) = 0$. Hence, $v_{\mathcal{M}^{\perp}{}_{FP}}(Par\ulcorner\tau\urcorner) = 0$ and so $Par\ulcorner\tau\urcorner$ is false at $v_{\mathcal{M}^{\perp}{}_{FP}}$. It follows that one can assert both $Par\ulcorner\lambda\urcorner$ and $\neg Par\ulcorner\tau\urcorner$.[37]

However, even with the modified falsity conditions it is not possible to correctly categorize every ungrounded unparadoxical statement. Let's take $\lambda \vee \tau$ as an example. This statement is not paradoxical according to Kripke's account—there are fixed-points where it is true. Yet, one will not be able to assert in the object-language that it is not paradoxical. The issue is that there are maximal fixed-points where this statement is $\frac{1}{2}$. So it will not be the case that for every fixed-point one can find a fixed-point extending it where the statement behaves bivalently. In other words, the statement is not semi-classical, and hence it is not semi-true, either. In fact, according to **Sophisticated Classification**, it is unstable. Thus, even under the amended characterization of paradoxicality, $v_{\mathcal{M}^{\perp}{}_{FP}}(Par\ulcorner\lambda \vee \tau\urcorner) = \frac{1}{2}$.[38]

Still, one may think that $\lambda \vee \tau$ is a very rare creature, with little philosophical significance. The thought would be that as long as one's theory is in a position to express that liars are paradoxical statements and truth-tellers are not, the theory has a claim to be adequate. In other words, one could suggest that from an explanatory point of view, the only conceptually important thought that a theory of paradoxicality should be able to capture is that liar-like statements are paradoxical but hypodoxical statements are not. Thus, one can basically ignore the problem posed by $\lambda \vee \tau$, since it belongs to neither category.

Unfortunately, this will not do. The theory is also unfit to correctly evaluate some hypodoxical statements as 'not paradoxical'. By way of example, consider again the statement $\tau_1 \vee (\lambda \wedge \tau_2)$ from above. There are fixed-points at which τ_1 is true. At those fixed-points $\tau_1 \vee (\lambda \wedge \tau_2)$ is true as well. There are also fixed-points at which both τ_1 and τ_2 are false. At those fixed-points, $\tau_1 \vee (\lambda \wedge \tau_2)$ is false. Therefore, $\tau_1 \vee (\lambda \wedge \tau_2)$ is hypodoxical. However, there are maximal fixed-points at which $\tau_1 \vee (\lambda \wedge \tau_2)$ is $\frac{1}{2}$. In particular, if τ_1 is false and τ_2 is true at some fixed-point, then $\tau_1 \vee (\lambda \wedge \tau_2)$ is $\frac{1}{2}$ at that fixed-point. Hence, $v_{\mathcal{M}^{\perp}{}_{FP}}(Par\ulcorner\tau_1 \vee (\lambda \wedge \tau_2)\urcorner) = \frac{1}{2}$. Once again, the

[37] There is another difference between the definitions that is worth highlighting. Consider a statement π saying of itself that it is not paradoxical, $\neg Par\ulcorner\pi\urcorner$. It is easy to check that $v_{\mathcal{M}_{FP}}(\pi) = \frac{1}{2}$, but $v_{\mathcal{M}^{\perp}{}_{FP}}(\pi) = 1$. Thus, the amended definition seems committed to the idea that there are statements that are true at a model purely in virtue of how they themselves semantically behave across the different fixed-points extending that model. This idea is incompatible with an intuition about truth that many would find plausible, namely, that the truth-value of a statement should ultimately depend on whether some non-semantic state of affairs obtains. For a discussion of the supervenience of semantics in fixed-point models, cf. Kremer (1988) and Gallovich (2022).

[38] Of course, Cook and Tourville are aware of this limitation.

characterization of paradoxicality is not as general as one might have expected it to be.

From the classical logician's perspective the obvious thing to do is to modify the falsity conditions for statements of the form $Par\ulcorner\phi\urcorner$ so as to make $Par(x)$ a bivalent predicate. One can stipulate that $v_{\mathcal{M}}(Par\ulcorner\phi\urcorner) = 0$ if it is not the case that $\forall\mathcal{M}' \in \mathcal{M}^{ext} : v_{\mathcal{M}'}(\phi) = \frac{1}{2}$. With this definition it holds that for every statement ϕ that is not paradoxical in Kripke's account, $Par\ulcorner\phi\urcorner$ will be false. But of course $Par(x)$ cannot behave bivalently in general without bringing paradoxes back.[39] To see this, consider again the Murzi-Rossi statement, ρ, saying of itself that it is paradoxical if true, $Tr\ulcorner\rho\urcorner \rightarrow Par\ulcorner\rho\urcorner$. With the bivalent definition, the construction does not reach a fixed-point for $Par(x)$. Of course, this ought not be surprising. One should not expect the paradoxicality predicate to behave bivalently in a paracomplete setting if this predicate is supposed to be untyped.

Yet, initially one might be tempted to impose the requirement that bivalence ought to be retained for statements belonging to the Par-free fragment of the language. After all, that is exactly the situation with Kripke's truth predicate. If ϕ belongs to the truth-free fragment of the language, then $Tr\ulcorner\phi\urcorner$ is either true or false. So one could hope that if ϕ does not contain occurrences of the paradoxicality predicate, then either $Par\ulcorner\phi\urcorner$ is true or it is false. Neither our characterization of $Par(x)$ nor Cook and Tourville's has this property. A natural question, then, is if it is possible to offer a definition of the paradoxicality predicate that respects this constraint.

At this point, we do not know. Once one adds $Tr(x)$ and $Par(x)$ to the language of PA some occurrences of Par can be implicit. That is, Par can occur as a part of a singular term that denotes (the code of) a statement that contains an implicit occurrence of Par, as in $Tr\ulcorner Par\ulcorner\psi\urcorner\urcorner$. But the problem is that it is not obvious that a suitable definition of 'implicit occurrence' can be given. One can construct statements of the form $\exists x(\phi(x) \wedge Tr(x))$, wherein $\phi(x)$ does not contain any explicit occurrences of Par but is true of $Par\ulcorner\psi\urcorner$ and only of that statement, i.e. it is provable in PA that $\forall x(\phi(x) \rightarrow x = \ulcorner Par\ulcorner\psi\urcorner\urcorner)$. Now, if one suggests that $\exists x(\phi(x) \wedge Tr(x))$ does not contain an implicit occurrence of Par, one has to deal with the fact that it is equivalent to $Par\ulcorner\psi\urcorner$ even though one contains an occurrence of Par and the other does not. If, instead, one suggests that $\exists x(\phi(x) \wedge Tr(x))$ does contain an implicit occurrence of Par (perhaps in virtue of its quantifying over a statement that contains an occurrence of it), then one needs to offer a suitable definition of 'statement that contains an implicit occurrence of Par'. However, that is a highly non-trivial task.[40]

At any rate, we do not think that this should be seen as a significant flaw of the approach. All in all, our account of paradoxicality is not altogether different from Kripke's account of truth. In the latter, the truth predicate is partial, so one must leave behind the (classical) expectation that every statement that fails to be true must be untrue. Similarly, if the paradoxicality predicate is partial, one must forgo the

[39] Unless, of course, one is willing to treat paradoxicality as a typed predicate. Indeed, it would be reasonable to expect a bivalent predicate if the goal were to extensionally capture Kripke's notion. But we have already noted that our goal is different, and so is Cook and Tourville's goal.

[40] One promising possibility is to offer a characterization of 'implicit occurrence' along the lines of Lavinia Picollo's account of alethic reference (Picollo 2015). However, pursuing this idea is beyond the scope of this paper.

expectation that every statement that fails to be paradoxical must be unparadoxical. Even though some statements that fail to be paradoxical cannot be said to be such in the object language, the characterization of the paradoxicality predicate that we have considered adequately captures the set of paradoxical statements in Kripke's theory, which is all that we set out to do. The predicate applies to a statement just in case the statement cannot obtain a classical truth-value. The main virtue of this analysis is not that it gets us one step closer to a universal language, but that it agrees with our intuitions about paradoxicality in a large number of cases.

5 Concluding remarks

According to the fixed-point conception of truth, the Tarskian hierarchy of languages fails to explain some aspects of our use of the notion of truth in natural language. As Kripke himself was ready to admit, the ghost of that hierarchy persists on his account, since it is only possible to talk about the status of the liar and of other statements from the metalanguage. Our account pushes the ghost further away by putting paradoxicality on a par with truth. We have offered a type-free theory of paradoxicality that is, to some extent, faithful to Kripke's original ideas. The theory allows us to talk about the semantic status of the liar and other paradoxical statements in the object-language. It also meets the desiderata that we have proposed for theories of paradoxicality in a natural and simple way. We find it reasonable to conclude that the fixed-point account has a lot going for it. It remains for future work to investigate what is the attractiveness of some of its competitors within the broader scenario of general conceptions of paradoxicality.

Acknowledgements Talks based on earlier versions of this paper were given (online) at the Work in Progress Seminar of the Buenos Aires Logic Group (Buenos Aires, 2021), at the workshop Truth in Expressively Rich Languages (Bristol, 2021), at the EXPRESS-PHILMAT Seminar (Amsterdam-Paris, 2021), and at the workshop The Philosophy of Metamathematical Results (Munich Center for Mathematical Philosophy, 2022). We wish to thank the organizers and participants of all these events. We are especially grateful to Eduardo Barrio, Aylén Bavosa Castro, Lwenn Bussiére-Caraes, Catrin Campbell-Moore, Bruno Da Ré, Hartry Field, Camillo Fiore, Daniel Grimaldi, Volker Halbach, Luca Incurvati, Hannes Leitgeb, Julien Murzi, Federico Pailos, Francesca Poggiolesi, Graham Priest, Ricardo Rodriguez, Ariel Roffé, Lorenzo Rossi, Mariela Rubín, Giorgio Sbardolini, Johannes Stern, Damián Szmuc, Paula Teijeiro, Omar Vasquez and Albert Visser. Last but not least, special thanks go to Luca Castaldo and two anonymous reviewers of Synthese for comments on previous drafts that led to significant improvements. Financial support for this work was provided by the project "Logic and Substructurality" (FFI2017-84805-P), funded by the Spanish MINECO (Ministerio de Economía, Industria y Competitividad).

References

Castaldo, L. (2021). Fixe-point Models for Paradoxical Predicates. *Australasian Journal of Logic, 18*(7), 688–723.

Cook, R. T. (2011). The No-No Paradox is a Paradox. *Australasian Journal of Philosophy, 89*(3), 467–482.

Cook, R. T. (2020). An intensional theory of truth: an informal report. *The Philosophical Forum, 51*(2), 115126.

Cook, R. T., & Tourville, N. (2016). Embracing the technicalities: expressive completeness and revenge. *The Review of Symbolic Logic, 9*(2), 325–358.

Cook, R. T., & Tourville, N. (2020). Embracing intensionality: paradoxicality and semi-truth operators in fixed point models. *Logic Journal of the IGPL, 28*(5), 747–770.
Eldridge-Smith, P. (2007). Paradoxes and hypodoxes of time travel. In P. W. Jan Lloyd Jones (Ed.), *Art and time* (pp. 172–189). Australia: Australian Scholarly Publishing.
Gallovich, C. (2022). *Semantic supervenience in fixed-point models*. Typescript.
Gallovich, C., & Rosenblatt, L. (2022). *Paradoxes, hypodoxes, and more*. Typescript.
Gilmore, P. C. (1974). The consistency of partial set theory without extensionality, axiomatic set theory. *Proceedings of Symposia in Pure Mathematics: American Math Society, 13*, 147–153.
Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic, 11*(1), 1–60.
Gupta, A. (2005). Do the paradoxes pose a special problem for deflationism? In J. Beall & B. Armour-Garb (Eds.), *Deflationism and paradox* (pp. 133–147). Oxford: Oxford University Press.
Hsiung, M. (2021). *In what sense is the no-no paradox a paradox? Philosophical studies*, forth- coming.
Kremer, M. (1988). Kripke and the logic of truth. *Journal of Philosophical Logic, 17*(3), 225–278.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*(19), 690–716.
Lawvere, W. (1969). Diagonal arguments and cartesian closed categories. *Reprints in Theory and Applications of Categories, 15*, 1–13.
Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophy Compass, 2*(2), 276–290.
Martin, R., & Woodruff, P. (1975). On representing 'True-in-L' in L. *Philosophia, 5*(3), 213–217.
Murzi, J., & Rossi, L. (2020). Generalized revenge. *Australasian Journal of Philosophy, 98*(1), 153–177.
Murzi, J., & Rossi, L. (2022). Non-reflexivity and revenge. *Journal of Philosophical Logic, 51*, 201–218.
Picollo, L. (2015). Minimalism, reference and paradoxes. In P. Arazim & M. Dancak (Eds.), *The logica yearbook 2015* (pp. 163–178). College Publications.
Prawitz, D. (1965). *Natural deduction: a proof-theoretical study*. New York: Dover.
Priest, G. (1994). The structure of the paradoxes of self-reference. *Mind, 103*(409), 25–34.
Priest, G. (2016). Logical disputes and the a priori. *Logique et Analyse, 236*, 347–366.
Rosenblatt, L. (2021). Paradoxicality without Paradox. Erkenntnis.
Rosenblatt, L., & Szmuc, D. (2014). On pathological truths. *The Review of Symbolic Logic, 7*(4), 601–617.
Rossi, L. (2019). A unified theory of truth and paradox. *The Review of Symbolic Logic, 12*(2), 209–254.
Tennant, N. (1982). Proof and paradox. *Dialectica, 36*, 265–296.
Walicky, M. (2017). Resolving infinitary paradoxes. *The Journal of Symbolic Logic, 82*(2), 709–723.