Paradoxes, Hypodoxes, and More

Camila Gallovich and Lucas Rosenblatt

Abstract Is it possible to provide a theory of truth that is capable of distinguishing the semantic status of paradoxical sentences from that of other ungrounded sentences without bringing meta-linguistic resources into play? We explore an account that extends Kripke's theory of truth with two primitive operators, one standing for the notion of paradoxicality and the other for the notion of hypodoxicality. Our results are mixed. While the paradoxicality operator behaves nicely, a number of restrictions need to be imposed to accommodate the hypodoxicality operator.

1 Introduction

A naive theory of truth endorses the claim that for every sentence ϕ, ϕ and *'ϕ' is true* are, in some sense, equivalent. Provided that self-referential sentences are available, this claim is incompatible with classical logic. In view of this, one might react in one of two ways, assuming, as we will do throughout the paper, that one is not willing to dramatically diminish the expressive resources of the language by banning self-referential sentences. Either one denies the principle stating that ϕ and *'ϕ' is true* are equivalent, or one eschews some principle of classical logic. In either case, the idea is not to reject the corresponding principle across the board, but rather to reject it for 'problematic' sentences only. The problematic sentences that we have in mind are sentences that talk about their own semantic status, like the liar sentence—a

Camila Gallovich
IIF-SADAF-CONICET, Bulnes 642, Buenos Aires, Argentina, e-mail: camila.gallovich@gmail.com

Lucas Rosenblatt
IIF-SADAF-CONICET, Bulnes 642, Buenos Aires, Argentina, e-mail: lucasrosenblatt83@gmail.com

sentence λ saying of itself that it is untrue—and the truth-teller sentence—a sentence τ saying of itself that it is true.

Of course, the liar sentence is problematic because it is responsible for the liar paradox. If one assumes that it is true, one can infer that it is false, and if one assumes that it is false, one can infer that it is true. That is, there is no consistent way of semantically evaluating it as true or as false. Other sentences, like the Curry sentence—that says of itself that if it is true, then 2+2=17 (or any other falsity)—are problematic for similar reasons. They are responsible for, in the sense that they contribute together with obvious logical and semantic principles to the production of, some paradoxical argument. However, some sentences are problematic in a very different way. For example, the puzzling nature of the truth-teller lies not in that it is responsible for a paradox, or in that there is no way of semantically evaluating it, but rather in that there are many ways of doing so, and no obvious reason to favor one over another.[1]

That the liar and the truth-teller are very different beasts is not only clear from a pre-theoretical standpoint. It is reflected, to some extent, in very influential accounts of the truth-theoretic paradoxes. For example, Saul Kripke cashes out the distinction between these sentences in terms of their behavior across a set of relevant interpretations.[2] Although, according to Kripke, both sentences are ungrounded (in a technical sense that we will make precise below), the liar is ungrounded because it cannot be in the extension or in the anti-extension of the truth predicate, whereas the truth-teller is ungrounded because one is free to consistently place it in the extension or in the anti-extension of the truth predicate.

For another example, the liar and the truth-teller will not behave in the same way across the sequences posited by the revision theory of truth.[3] The liar is semantically unstable—it goes in and out of the extension of the truth predicate—in a way that the truth-teller is not. One is at liberty to place the truth-teller either in the extension or in the anti-extension of the truth predicate at the start of a revision sequence. But once it is placed in one or in the other, it stays there for good, so to speak. Other approaches similarly distinguish these sentences on account of their divergent semantic behaviour.

This leads us to our main point. This difference is seldom, if ever, reflected in the object language treatment of the corresponding sentences. Some accounts settle for a meta-linguistic diagnosis, that is, they recognize that the distinction between different problematic sentences is important from the meta-language, but they fail to make it explicit in any way in the object language. Other accounts do not settle for a meta-linguistic diagnosis, and they sometimes even introduce a sentence-classifying

[1] The distinction extends straightforwardly to sets of sentences, such as the paradoxical 'postcard pair'—consisting of a pair of sentences λ_1 saying of λ_2 that it is untrue, and λ_2 saying of λ_1 that it is true—and the non-paradoxical 'open pair'—consisting of two sentences λ_a and λ_b saying of the other that it is untrue. We should note, though, that the unparadoxicality of the open pair is a matter of some controversy. Cf. [29] and [7].

[2] This is done in his classic essay on truth, [19].

[3] The loci classici are Hans Herzberger's paper [18], Anil Gupta's paper [16], and Nuel Belnap & Anil Gupta's book [17].

or 'diagnostic' operator to the object language specifically designed for the purpose of semantically categorizing the problematic sentences.[4] Yet, for all their merit, these accounts are unsatisfactory for the task at hand. Although in them it is possible to use this operator to say (in the object language) of the problematic sentences that they are problematic, the operator typically lumps all these sentences together, completely disregarding the distinction that we are aiming at. That is, one is able to express that the liar and other paradoxical sentences are problematic, and also that the truth-teller and its ilk are problematic, but that's it. There is no way of expressing that the liar is problematic in a different way that the truth-teller is. On account of this, we think that a more fine-grained treatment is called for.[5]

In what follows we will explore one way in which this can be done. The main goal of the paper, then, is to offer an account of truth that treats the liar and its ilk differently from the truth-teller and its ilk, and that does so without bringing into play meta-linguistic resources not available in the object-language. The paper is structured as follows. In the next section we present Kripke's fixed-point approach and we show how to extend it with a couple of operators, one standing for paradoxicality and the other for hypodoxicality (on which more shortly). After that, we discuss a number of features of this enriched Kripkean approach. In Section 3 we examine a few objections, open problems and limitations. In Section 4 we offer some concluding remarks.

2 Paradoxicality and hypodoxicality made precise

2.1 Kripke's approach and two projects

Undoubtedly, one of the most influential treatments of the truth-theoretic paradoxes is given by the fixed-point approach developed by Saul Kripke in his *Outline of a Theory of Truth*. An interesting aspect of Kripke's theory that is often neglected is that it can be viewed as providing a meta-linguistic characterization of the sentences of a certain

---

[4] Several approaches fitting this description can be found in the literature. Hartry Field's works on determinacy (cf. e.g. [12]), J.C. Beall's paranormal-based approach (cf. [2]), Lucas Rosenblatt and Damián Szmuc's theory of pathologicality (cf. [28]), and Andrew Bacon's recent healthiness account (see [1]). The list could go on.

[5] An exception to this (i.e., to the lumping together of problematic sentences) can be found in Roy Cook and Nicholas Tourville's recent paper [10] and in Cook's solo paper [8]. Cook and Tourville's goal is, roughly, to obtain a framework that avoids revenge paradoxes by being capable of expressing every semantic concept, even those that, on the face of it, are intensional, like paradoxicality. Although our goal here is slightly different, some of the ideas that we will explore in this paper, as well as some of the problems that we will discuss, have been very much inspired by Cook and Tourville's work (cf. also their earlier paper [9]). As for Cook's solo paper, his purpose is to provide a systematic study of intensional operators within the fixed-point approach. The main philosophical idea underpinning his account—namely, that a theory of truth capable of making intensional distinctions is desirable—is one that we are completely on board with. We'll come back to this in Section 3.3.

formal language in terms of their semantic behaviour. More specifically, according to Kripke's theory, there are two types of sentences, those that are grounded and those that are not. A *grounded* sentence is one that obtains a classical truth-value (true or false) at the minimal fixed-point.[6] Ungrounded sentences come in many varieties. We'll say, following Kripke, that an ungrounded sentence is *paradoxical* if it lacks a truth-value at every fixed-point. That is, it is not possible to assign a (classical) truth-value to a paradoxical sentence at a fixed-point, which means that, at every fixed-point, the sentence is neither true nor false.[7]

But not every ungrounded sentence is paradoxical. Let's say that an ungrounded sentence is *hypodoxical* if there are at least two fixed-points that evaluate it differently, one that makes it true and one that makes it false.[8] Grounded sentences, like 'two plus two equals four', and paradoxical sentences, like the liar, have something in common. They present a constant behaviour across the fixed-points. The difference is that while grounded sentences always have a classical value, paradoxical sentences never do. Hypodoxical sentences do *not* behave constantly across the set of fixed-points, so they are neither grounded nor paradoxical. An example of this is given by the truth-teller, which is true at some fixed-points and false at other fixed-points.[9]

These notions are an essential part of Kripke's account. Yet, there's no way of expressing them in the object language of his theory. An example that is often mentioned in connection to this is that in Kripke's account one cannot say anything about the semantic status of the liar sentence. In particular, one cannot say that it is neither true nor false. From a technical viewpoint, this amounts to the impossibility of introducing a (bivalent, truth-functional) 'neither true nor false' operator to the object language without bringing about new paradoxes. This is sometimes identified in the literature as an example of a revenge paradox.

Here we are not so much interested in revenge paradoxes—at least not as they are usually understood—but in a somewhat more subtle phenomenon, that of separating paradoxicality and hypodoxicality without resorting to the meta-language. The difference, as we see it, is that if one tries to reflect the distinction between paradoxes and hypodoxes in the object language one does not necessarily end up with new paradoxes. Rather, it is simply that most theorists have typically neglected the

---

[6] In what follows we will assume that the reader is familiar with the basic concepts of Kripke's theory.

[7] Traditionally, paradoxicality is taken to be a property that applies primarily to arguments, not to sentences. Here, however, our focus will be on sentences, since we are dealing with paradoxicality *within Kripke's theory of truth*. Although the question of which are the primary bearers of paradoxicality is interesting, it is beyond the scope of this paper.

[8] To the best of our knowledge, the term 'hypodox' was coined by Peter Eldridge-Smith in [11] (see fn. 3 of that paper) to refer to sentences like the truth-teller, pairs of sentences like the open pair, and so on, such that, in his words, "might consistently take either truth-value but we have no basis for determining which" (p. 178).

[9] A very radical view about hypodoxes has been recently advanced by Alexandre Billon in [4]. He argues that hypodoxes, such as the truth-teller, are just as paradoxical as the liar. For reasons of space, we cannot engage with this claim here. We will take for granted that the truth-teller and its ilk are *not* paradoxical, or, at least, that if they are paradoxical, they are so in a very different way that the liar and its ilk are.

distinction, or have underestimated the importance of drawing it without bringing metalinguistic resources into play.[10] Yet, we believe there are a couple of reasons for thinking that these concepts ought to receive an object-linguistic treatment.

One of the reasons is related to what J.C. Beall calls the 'exhaustive characterization project'.[11] In his words, this project consists in

> explain[ing] how, if at all, we can truly characterize—specify the 'semantic status' of—all sentences of our language (in our language). [3, p. 66]

Even if one is not fully on board with the project, it seems that it would be desirable to characterize the semantic status of at least some paradigmatic problematic sentences of the language without resorting to the metalanguage. However, it is usually left unspecified how this categorization ought to be carried out. Certainly, most approaches that are said to satisfy the constraint imposed by the exhaustive characterization project do manage to categorize every sentence. The problem, we contend, is that the characterization is not fined-grained enough, and that this is so quite independently of the notion being chosen to perform it.

Let us illustrate why this is so. A usual strategy to fulfill the exhaustive characterization project is to introduce some unary diagnostic operator that applies to sentences and that, intuitively, attaches to them some meta-linguistic property on the basis of their semantic behavior. If one defines this operator in the right manner, it should deliver a way of distinguishing in the object language sentences that are 'well-behaved' from sentences that are 'badly-behaved'. For example, in Kripke's approach one could introduce a groundedness operator. If an appropriate operator of this sort is available, one will be able to separate the sentences that are grounded from the sentences that are not, but one will fail to distinguish ungrounded sentences that are paradoxical from ungrounded sentences that aren't so. If, instead, a paradoxicality operator is introduced, one will not be able to tell the difference between unparadoxical sentences that are grounded and unparadoxical sentences that are ungrounded. If, lastly, a hypodoxicality operator is made available, one will not be able to segregate unhypodoxical sentences that are grounded from unhypodoxical sentences that are paradoxical.[12] In all these cases we can meta-linguistically recognize at least three types of sentences, but the corresponding operator only splits sentences

[10] Although see Murzi and Rossi's [20] for an argument to the effect that introducing a 'naive' paradoxicality predicate in the object language of Kripke's approach generates new paradoxes. For a discussion of their argument, cf. [26] and [27].

[11] This may also be related to what Charles Chihara refers to as the 'diagnostic problem', since, as theorists, what we are after is not simply a solution that identifies what went wrong in certain paradoxical arguments, but a theory that is capable of providing an informative diagnosis (or characterization) of how the 'diseased' sentences produce those paradoxes. In Chihara's words [6, p. 590], one should be able to explain "(...) how and why the deception was produced (...)".

[12] Other choices of operators, such as ungroundedness, unparadoxicality, and unhypodoxicality, fall prey to the same difficulty.

into two categories. Unavoidably, then, some sentences are lumped together when they shouldn't be. So, at best, the characterization is only partially successful.[13]

In order to offer a semantic characterization of liars and truth-tellers that avoids their conflation, the use of two different operators appears to be indispensable. And although there is more than one way of doing this, our strategy will be to avail ourselves of both a paradoxicality operator and a hypodoxicality operator. If one has the expressive resources to talk about paradoxicality and hypodoxicality—and not just about the more general notion of ungroundedness—one will be in a position to preempt lumping all ungrounded sentences together.[14]

In our approach there will be a trade-off between fine-grainedness and expressive completeness. As we see it, the importance of semantically characterizing each sentence of the language in the language is sometimes overstated. We think that it is more fundamental to offer a categorization—even one that is not exhaustive and thus remains silent on the status of certain sentences—that makes all the required distinctions, or at least most of them. For this the presence of two operators is necessary.

The second reason for introducing object-language diagnostic operators has to do with a different albeit related project, which usually goes by the name of 'classical recapture'. Typically, non-classical approaches to the truth-theoretic paradoxes do not reject classical reasoning in every context. Even if the underlying logic denies the validity of some classical principles, the non-classical theorist insists on retaining (instances of) these principles in certain contexts. Thus, there is an onus on her to provide a rigorous method to distinguish the contexts where it is 'safe' to reason classically from the contexts where it is not.[15] Given that this distinction plays such an important explanatory role in these approaches, arguably one ought to have the resources to draw it in the object language. The presence of diagnostic operators can help with that too.[16]

2.2 Kleene-Kripke interpretations

Before presenting our proposal we need to set the formal stage and to briefly discuss some technical preliminaries. We will work with a quantifier-free language $\mathcal{L}$. We'll

[13] We hasten to add that we are not claiming that this (very demanding) understanding of the exhaustive characterization project is entirely faithful to Beall's original motivations, and we don't know whether he would endorse it—probably not.

[14] A note of warning is necessary at this point. Because of the way in which we've defined hypodoxicality, even the presence of these two operators is insufficient to classify all ungrounded sentences appropriately. In particular, there are ungrounded sentences that are neither paradoxical nor hypodoxical. We will come back to this in Section 3.3.

[15] Fine-grainedness is also important here, for the aforementioned problem of lumping ungrounded sentences together might prevent us from distinguishing (in the object language) safe contexts—say, those involving grounded sentences—from potentially dangerous contexts—e.g. those involving paradoxical sentences.

[16] For more on the idea of classical recapture, cf. [24] and [25].

assume that the primitive logical connectives of $\mathcal{L}$ are negation, $\neg$, conjunction, $\wedge$, and a truth constant, $\top$. Other logical connectives such as disjunction, $\vee$, and the conditional, $\rightarrow$, can be defined in terms of them. The extra-logical vocabulary of $\mathcal{L}$ contains an infinite number of propositional variables p_1, p_2, p_3, ... We will say that $\mathcal{L}^{Tr}$ is the result of adding to $\mathcal{L}$ a one-place predicate, $Tr(x)$, intended to represent truth, and an infinite set of constants $A = \{a_1, a_2, a_3, ...\}$. The denotation of each of the constants a_i in A is given by a one-to-one function f such that $f : A \rightarrow Sent_{\mathcal{L}^{Tr}}$, where, as usual, $Sent_{\mathcal{L}^{Tr}}$ stands for the set of $\mathcal{L}^{Tr}$-sentences. For instance, depending on one's choice of f, there can be a constant a_j such that, according to f, its denotation is the sentence $\neg Tr(a_j)$, i.e. a liar sentence. Also, there can be a constant a_k such that, according to f, its denotation is the sentence $Tr(a_k)$, i.e. a truth-teller. Thus, the distinction between paradoxical and hypodoxical sentences can already be cast in this very simple framework.

Since our goal is to give an account of the notions of hypodoxicality and paradoxicality within the object language, we'll add two unary operators, $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$ (standing for hypodoxicality and paradoxicality, respectively) to $\mathcal{L}^{Tr}$, and another set of constants $B = \{b_1, b_2, b_3, ...\}$.[17] The resulting language will be called $\mathcal{L}^+$. The denotation of each of these new constants is given by another one-to-one function g such that $g : A \cup B \rightarrow Sent_{\mathcal{L}^+}$, where $Sent_{\mathcal{L}^+}$ stands for the set of $\mathcal{L}^+$-sentences, and such that for each a_i, g agrees with f, that is, $g(a_i) = f(a_i)$. For example, depending on one's choice of g, there can be a constant b_j such that, according to g, its denotation is the sentence $O^{\mathcal{P}}Tr(b_j)$, i.e. a sentence saying of itself that it is paradoxical. Also, there can be a constant b_k such that, according to g, its denotation is the sentence $O^{\mathcal{H}}Tr(b_k)$, i.e. a sentence saying of itself that it is hypodoxical. Throughout the paper we will work with a fixed denotation function g and we will assume that g is such that (i) for every b in B its denotation according to g is not in $\mathcal{L}^{Tr}$, and (ii) it interprets the constants a_j, a_k, b_j and b_k in the way suggested. To make the notation more perspicuous, we will use corner quotes $\ulcorner\urcorner$ to name sentences. More specifically, one can think of $\ulcorner\urcorner$ as a name forming device such that $\ulcorner\phi\urcorner$ is a name for the sentence ϕ.[18]

To semantically interpret the language $\mathcal{L}^+$ we will rely on strong Kleene interpretations.[19] A *strong Kleene interpretation* for $\mathcal{L}^+$ is a function, v, that assigns a semantic value in $\{1, \frac{1}{2}, 0\}$ to the sentences of $\mathcal{L}^+$.

---

[17] We could, alternatively, directly employ predicates for paradoxicality and hypodoxicality. The choice is mostly a matter of taste, as these predicates can be unproblematically emulated using the truth predicate together with the corresponding operators. That is, in our framework a paradoxicality predicate, $Par(x)$, can be explicitly defined as $O^{\mathcal{P}}Tr(x)$, and a hypodoxicality predicate, $Hyp(x)$, can be explicitly defined as $O^{\mathcal{H}}Tr(x)$.

[18] We've lifted the idea of constructing self-referential sentences in this way from Dave Ripley's [22]. The method has also been used in various other places, e.g. [17]. It would be natural for the reader to ask why we are employing a quantifier-free language and this non-standard method of modelling (self-)reference. After all, typically, if one wishes to discuss self-referential sentences, the obvious choice is to use first-order arithmetic as one's base theory, or some other theory expressive enough to code facts about syntax. The answer will have to wait until Section 3.4.

[19] Kripke's approach is also compatible with other kinds of interpretations, such as those based on the weak Kleene schema or various types of supervaluational schemata. We think that most of the

Definition 1 The function v satisfies the following conditions:

- For propositional variables p, $v(p) \in \{1, 0\}$.[20]
- $v(\top) = 1$.
- $v(\neg\phi) = 1 - v(\phi)$.
- $v(\phi \wedge \psi) = min(v(\phi), v(\psi))$.[21]

If, in addition, the interpretation satisfies the constraint below, we will say that it is a *Kleene-Kripke interpretation* (or KK-*interpretation*, for short):

- $v(Tr(b)) = v(g(b))$.

KK-interpretations are fixed-points in Kripke's sense, so if an interpretation v is a KK-interpretation, we will also say that v is a *Kripke fixed-point*.[22]

Kripke's original construction uses the following notion of extension:

Definition 2 (*Extension*)
A strong Kleene interpretation v' *extends* a strong Kleene interpretation v if and only if for all ϕ in $\mathcal{L}^+$:

- if $v(\phi) = 1$, then $v'(\phi) = 1$, and
- if $v(\phi) = 0$, then $v'(\phi) = 0$.

Informally, if an interpretation v' (properly) extends an interpretation v, then v' is more informative than v in that v is silent about the truth-value of some sentences (in the sense that it declares them neither true nor false) that v' declares either true or false. The relation of one interpretation extending another allows us to compare different interpretations in a very useful way. In Kripke's construction, we can start with an interpretation where every sentence of the form $Tr\ulcorner\phi\urcorner$ is neither true nor false. Then, if ϕ is true (false) at that interpretation, we can construct another interpretation that extends it and that makes $Tr\ulcorner\phi\urcorner$ true (false). One proceeds in this way for a long time, through the finite ordinals and many infinite ordinals, obtaining larger and larger interpretations, each extending all the ones that came before. At some stage, a (Kripke) fixed-point is reached. In a (very tight) nutshell, this is how Kripke's construction works.

Here's one natural proposal to interpret the new operators using the notion of extension. Intuitively, the sentence $O^{\mathcal{P}}\phi$ expresses that ϕ is paradoxical and the sentence $O^{\mathcal{H}}\phi$ expresses that ϕ is hypodoxical. We can say that a sentence $O^{\mathcal{P}}\phi$ is true at an interpretation if and only if for every Kripke fixed-point extending that interpretation, ϕ is neither true nor false. And we can say that a sentence $O^{\mathcal{H}}\phi$ is true at an interpretation if and only if there is a Kripke fixed-point extending that interpretation where ϕ is true and there is a Kripke fixed-point extending that

---

things we say below could be adapted to these other frameworks, but a careful analysis of this is beyond the scope of this paper.

[20] This establishes that sentences belonging to the base language, $\mathcal{L}$, behave bivalently.

[21] $min(v(\phi), v(\psi))$ is the operation that outputs the minimum of $v(\phi)$ and $v(\psi)$.

[22] For the details of Kripke's construction we refer the reader to [14].

interpretation where ϕ is false. These are straightforward generalizations of Kripke's original notions. For the paradoxicality operator the generalization works fine, and it is also satisfactory for *certain sentences* involving the hypodoxicality operator. For example, sentences like $O^{\mathcal{H}}\phi$, where ϕ is grounded, will be false, and sentences like $O^{\mathcal{H}}\tau$ will be true. But it dramatically fails for sentences containing iterations of $O^{\mathcal{H}}$, such as $O^{\mathcal{H}}O^{\mathcal{H}}\phi$. The definition will make all sentences of this kind true, given that if one starts with an interpretation where all sentences of the form $O^{\mathcal{H}}\phi$ are neither true nor false—which is the only natural decision to make concerning sentences that haven't been interpreted yet—there will be an interpretation extending this initial interpretation where this sentence is true, and another interpretation also extending this initial interpretation where this sentence is false. But then $O^{\mathcal{H}}O^{\mathcal{H}}\phi$ will be true, regardless of the truth-value of ϕ. For this reason, we need to slightly modify the characterization of the hypodoxicality operator.

In order to do this, the sentences of the language must be separated into two sets. The first set features sentences that do not contain occurrences of $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$, and the second set features sentences that do contain such occurrences. Note that some occurrences could be implicit, in the sense that the operator does not occur as part of a subformula of the corresponding sentence, but it occurs as part of a sentence that is denoted by a name occurring in the sentence. For example, in the sentence $O^{\mathcal{P}}\lambda$ the operator $O^{\mathcal{P}}$ occurs explicitly, whereas in $Tr\ulcorner O^{\mathcal{P}}\lambda\urcorner$ it occurs implicitly. Observe that implicit occurrences of an operator can occur arbitrarily deep. For instance, $O^{\mathcal{P}}$ also occurs implicitly in $Tr\ulcorner Tr\ulcorner O^{\mathcal{P}}\lambda\urcorner\urcorner$. The sets of constants A and B given above for $\mathcal{L}^{Tr}$ and $\mathcal{L}^{+}$, respectively, allow us to draw those distinctions rigorously.

Definition 3 (*Explicit and Implicit Occurrences*)
A sentence ϕ contains an explicit occurrence of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$ if and only if ϕ contains a sub-formula of the form $O^{\mathcal{P}}\psi$ or $O^{\mathcal{H}}\psi$, respectively. A sentence ϕ contains an implicit occurrence of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$ if and only if ϕ contains a constant b that belongs to the set B. We will say that a sentence ϕ contains an occurrence of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$ if ϕ contains an implicit or explicit occurrence of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$, respectively.[23]

Recall that the constants belonging to B only denote sentences that contain (explicitly or implicitly) some occurrences of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$, whereas the constants belonging to A don't. It follows that a sentence contains an occurrence of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$ if and only if it belongs to $\mathcal{L}^{+}$ but not to $\mathcal{L}^{Tr}$. So, given a sentence ϕ, either one of the operators occurs in it or it doesn't. In the first case, the sentence belongs to $\mathcal{L}^{+}$ but not to $\mathcal{L}^{Tr}$. If the operator occurs implicitly in the sentence, then it will contain some name b belonging to B.

Using these facts we can provide an alternative characterization of hypodoxicality that avoids the previous difficulty. We will say that a sentence ϕ is hypodoxical if and only if there is an interpretation at which it is true, another interpretation at which it is false, and ϕ belongs to $\mathcal{L}^{Tr}$ (i.e. ϕ doesn't contain any occurrences of one of the new operators). Below we make this precise.

[23] Of course, the definition assumes a fixed denotation function g for the members of the set B.

2.3 The construction

Showing how to adequately interpret a language containing $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$ is anything but trivial. Once these operators are available, one can construct sentences that attribute paradoxicality or hypodoxicality to sentences of the base language, but also to other sentences that in turn contain occurrences of $O^{\mathcal{P}}$, $O^{\mathcal{H}}$, or both. Furthermore, because it is possible to iterate these operators, it is also possible to construct sentences that assert of themselves that they are paradoxical or hypodoxical. Our theory should explain all these cases.

A bit more generally, we submit that there are a number of reasonable desiderata that a Kripkean theory of paradoxicality and hypodoxicality ought to satisfy. Here we will focus on three. First, the theory should be capable of separating liars from truth-tellers. In particular, it should assert that liars are paradoxical and that truth-tellers are hypodoxical, but it should *not* assert that liars are hypodoxical or that truth-tellers are paradoxical. Secondly, the theory has to agree with Kripke's original account in the following sense: if a sentence is paradoxical according to Kripke's account, then the theory should assert that it is, and if a sentence is hypodoxical according to Kripke's account, the theory should assert that it is. Third, the new operators should interact nicely with the rest of the language. For example, the theory should validate inferences such as "$O^{\mathcal{H}}\phi$ implies $O^{\mathcal{H}}\neg\phi$".[24]

Here's an informal explanation of how the operators $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$ will be interpreted. The general idea is to set up a *macro*-construction for $O^{\mathcal{H}}$, such that at each stage one runs a *micro*-construction for $O^{\mathcal{P}}$ (and, although we won't bother to spell this out, such that at each stage of the micro-construction one runs a *sub*-construction for $Tr(x)$ which is, basically, Kripke's fixed-point construction).[25] We will call the entire construction, $\mathfrak{C}$.

$\mathfrak{C}$ starts with an interpretation $v^{<0,0>}$ which is a minimal fixed-point of the Kripke construction. In $v^{<0,0>}$ every grounded sentence of $\mathcal{L}^{Tr}$ is *either* true or false, and all sentences of the form $O^{\mathcal{P}}\phi$ or of the form $O^{\mathcal{H}}\phi$ are *neither* true nor false. Note that $v^{<0,0>}$ is just one among the many possible starting interpretations that could kick off the construction $\mathfrak{C}$: we can take any bivalent interpretation for $\mathcal{L}$ and obtain a minimal fixed-point using Kripke's technique. Thus, one should bare in mind that $\mathfrak{C}$ is always relativized to an interpretation for $\mathcal{L}$, and so there are as many possible starting interpretations for $\mathfrak{C}$ as there are minimal fixed-points obtainable from interpretations for the base language $\mathcal{L}$.[26]

One then proceeds in stages. This has to be done sequentially, so we first interpret sentences of the form $O^{\mathcal{P}}\phi$, and only then interpret sentences of the form $O^{\mathcal{H}}\phi$. In

[24] There are various other desiderata that one could impose, but we have decided to focus on these three because they strike us as particularly important. For a more detailed discussion of this, see [27].

[25] There is no particular reason to interpret sentences of the form $O^{\mathcal{P}}\phi$ first and sentences of the form $O^{\mathcal{H}}\phi$ second. But some order needs to be chosen. As far as we can see, there are no difficulties if the order is reversed. The idea of employing a macro-construction and a micro-construction is taken from Hartry Field's [13], although he uses them to interpret two different conditionals.

[26] This will be important later on, when we discuss contingent paradoxes (see Section 3.2).

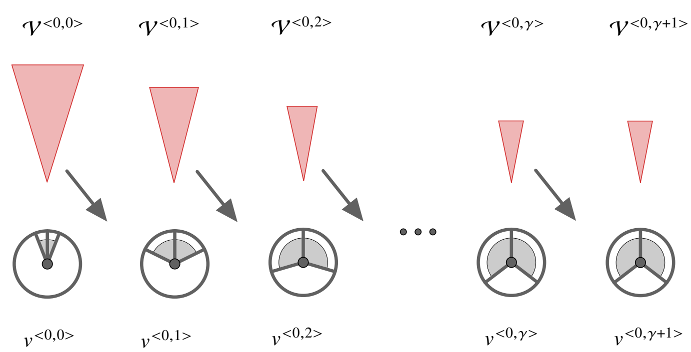
other words, we first run the micro-construction for $O^{\mathcal{P}}$. This yields a sequence of interpretations

$$v^{<0,0>}, v^{<0,1>}, v^{<0,2>}, \ldots, v^{<0,\gamma>}, v^{<0,\gamma+1>}, \ldots$$

To calculate the truth-value of sentences of the form $O^{\mathcal{P}}\phi$ at any of these interpretations, we need to look at all the different ways of interpreting the sentence ϕ at the prior stage. Note that every interpretation $v^{<\alpha,\beta>}$ has an associated set whose members are all the Kripke fixed-points that extend it—we call this set $\mathcal{V}^{<\alpha,\beta>}$.

$$\mathcal{V}^{<0,0>}, \mathcal{V}^{<0,1>}, \mathcal{V}^{<0,2>}, \ldots, \mathcal{V}^{<0,\gamma>}, \mathcal{V}^{<0,\gamma+1>}, \ldots$$

For example, in order to determine the truth-value of a sentence $O^{\mathcal{P}}\phi$ at the first stage of the first micro-construction, $v^{<0,1>}$, we look at the set $\mathcal{V}^{<0,0>}$, which is the set of all KK-interpretations extending $v^{<0,0>}$. The set $\mathcal{V}^{<0,0>}$ is used to interpret sentences of the form $O^{\mathcal{P}}\phi$, as their value will depend on the behaviour of ϕ across the KK-interpretations in $\mathcal{V}^{<0,0>}$. Once we've determined the truth-value of these sentences, we calculate the truth-values of the rest of the sentences of $\mathcal{L}^+$ using the KK-operations. That is, after the truth-values of the sentences of the form $O^{\mathcal{P}}\phi$ are settled, we can carry out the usual Kripke fixed-point construction. The different KK-interpretations that can be reached in this way are now used to calculate the truth-values of the sentences of the form $O^{\mathcal{P}}\phi$ at the next stage. After that, we calculate the truth-value of the rest of the sentences of $\mathcal{L}^+$ using the KK-operations again. When we're done with that, we move to the next stage. This process is repeated at every successor stage $\alpha + 1$. At limit stages we look at the intersection of the previous interpretations (details are provided below). At some stage we reach a fixed-point for $O^{\mathcal{P}}$, $v^{<0,\gamma>} = v^{<0,\gamma+1>}$. The procedure can be represented as follows:

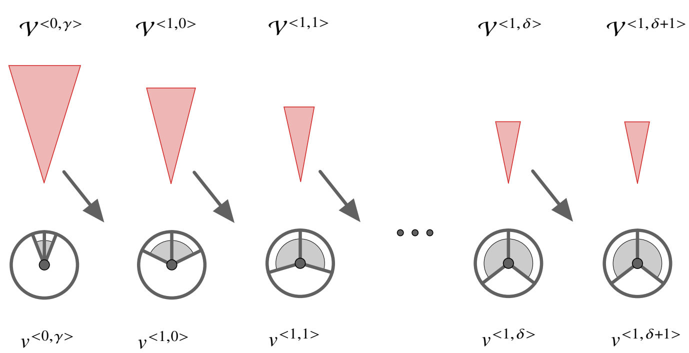


We say that $v^{<0,\gamma>}$ is a fixed-point for $O^{\mathcal{P}}$ because in order to determine the truth-value of sentences of the form $O^{\mathcal{P}}\phi$ at $v^{<0,\gamma>}$ we need to look at the behavior of ϕ at the set of KK-interpretations $\mathcal{V}^{<0,\gamma>}$ extending that very same interpretation.[27]

We then use $\mathcal{V}^{<0,\gamma>}$ to obtain the first interpretation of the second micro-construction, $v^{<1,0>}$. At this interpretation some sentences of the form $O^{\mathcal{H}}\phi$ obtain a value different from $\frac{1}{2}$. Then, starting from $v^{<1,0>}$, we run the micro-construction again, obtaining a new sequence of interpretations

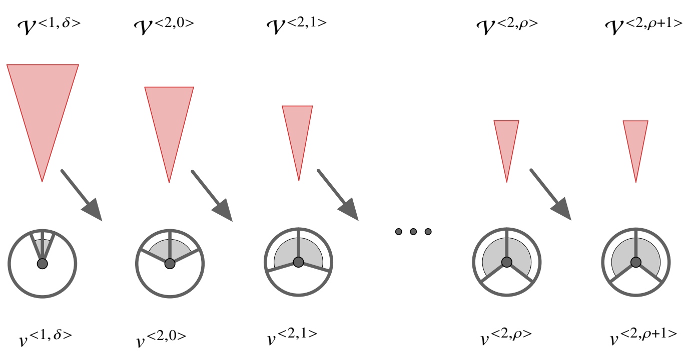
$$v^{<1,0>}, v^{<1,1>}, v^{<1,2>}, \ldots, v^{<1,\delta>}, v^{<1,\delta+1>}, \ldots$$

This is done until we reach a new fixed-point for $O^{\mathcal{P}}$, $v^{<1,\delta>} = v^{<1,\delta+1>}$. This part of the construction can be represented as follows:



We then use $\mathcal{V}^{<1,\delta>}$ to obtain the first interpretation of the third micro-construction, $v^{<2,0>}$. This, in turn, is the starting point of another sequence, as the next picture shows:

[27] Because we've restricted ourselves to a quantifier-free language, for each α, the closure ordinal of the micro-construction starting from $v^{<\alpha,0>}$ is ω. So the micro-construction reaches a fixed-point for $O^{\mathcal{P}}$ at $v^{<\alpha,\omega>}$. But of course this is just a consequence of using $\mathcal{L}^{+}$. For richer languages the closure ordinal could be different and we want the picture to represent that as well, so that's why we've used '$v^{<0,\gamma>}$' for the fixed-point. The pictures below are misleading in other ways too, but we've decided to keep them because we think that they are helpful in spite of these inaccuracies.



And so on and so forth. Every time one of the micro-constructions reaches a fixed-point for $O^{\mathcal{P}}$, we start a new micro-construction. If we only look at the starting points of each micro-construction, we have the following sequence of KK-interpretations:

$$v^{<0,0>}, \ldots, v^{<1,0>}, \ldots, v^{<2,0>}, \ldots, v^{<3,0>}, \ldots, v^{<\mu,0>}, \ldots, v^{<\mu+1,0>}, \ldots$$

At some stage we reach a fixed-point $v^{<\mu,0>} = v^{<\mu+1,0>}$ for the entire macro-construction, which will be a fixed-point for both $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$. Below, all of this will be made precise.

The construction we are about to offer is heavily based on a technique introduced, as far as we know, in Stephen Yablo's [31]. The account resembles a possible world semantics, in that KK-interpretations play the role of possible worlds and the extension relation plays the role of the accessibility relation. One evaluates attributions of paradoxicality and hypodoxicality at some KK-interpretation by considering KK-interpretations different from it.[28]

With this in mind, it is now time to provide a more rigorous characterization of the construction $\mathfrak{C}$ and of how the paradoxicality and hypodoxicality operators behave in it.

How $\mathfrak{C}$ starts: $v^{<0,0>}$ and $\mathcal{V}^{<0,0>}$.

We define $v^{<0,0>}$ as the minimal Kripke fixed point[29] (where all sentences of the

[28] The idea of applying Yablo's technique to provide a model-theoretic treatment of a sentence-classifying operator ('pathologicality') is proposed and explored in a paper written by one of us (LR, together with Damián Szmuc), [28]. As Cook and Tourville point out in [10], the construction on offer there can be simplified in several ways. In particular, from a technical point of view, there was no need to go beyond a three-valued semantics (the models used in [28] are penta-valued). We have tried to follow their suggestion here.

[29] As we noted at the beginning of this subsection, one should always bare in mind that the construction has an initial valuation v as a parameter. That is, the interpretation $v^{<0,0>}$ should always be understood as relativized to an initial classical valuation v for the base language $\mathcal{L}$.

form $O^{\mathcal{P}}\phi$ and of the form $O^{\mathcal{H}}\phi$ have the value $\frac{1}{2}$) and we characterize $\mathcal{V}^{<0,0>}$ in the following way:

$$\mathcal{V}^{<0,0>} = \{v : v \text{ is a KK-interpretation extending } v^{<0,0>}\}$$

Successor ordinals for $O^{\mathcal{P}}$: $v^{<\alpha,\beta+1>}$ and $\mathcal{V}^{<\alpha,\beta+1>}$.

For each ordinal α, let $v^{<\alpha,\beta+1>}$ be the interpretation obtained by letting sentences of the form $O^{\mathcal{P}}\phi$ behave as specified below and then applying the KK-operations:

$$v^{<\alpha,\beta+1>}(O^{\mathcal{P}}\phi) = \begin{cases} 1 & \text{if } \forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = \frac{1}{2} \\ 0 & \text{if } \forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = \text{i, for i} \in \{1,0\}^{30} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

$$\mathcal{V}^{<\alpha,\beta+1>} = \{v : v \in \mathcal{V}^{<\alpha,\beta>} \text{ and } v \text{ is a KK-interpretation extending } v^{<\alpha,\beta+1>}\}$$

Limit ordinals for $O^{\mathcal{P}}$: $v^{<\alpha,\lambda>}$ and $\mathcal{V}^{<\alpha,\lambda>}$.

For each ordinal α, let $v^{<\alpha,\lambda>}$ (where λ a limit ordinal, not the liar sentence!) be the interpretation obtained by letting sentences of the form $O^{\mathcal{P}}\phi$ behave as specified below and then applying the KK-operations:

$$v^{<\alpha,\lambda>}(O^{\mathcal{P}}\phi) = \begin{cases} 1 & \text{if } \forall v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \beta < \lambda\} : v(\phi) = \frac{1}{2} \\ 0 & \text{if } \forall v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \beta < \lambda\} : v(\phi) = \text{i, for i} \in \{1,0\} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

$$\mathcal{V}^{<\alpha,\lambda>} = \{v : v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \beta < \lambda\} \text{ and } v \text{ is a KK-interpretation extending } v^{<\alpha,\lambda>}\}$$

If a KK-interpretation v satisfies the clause above, we will say that v is a *fixed-point for $O^{\mathcal{P}}$*.[31]

[30] This should be understood as: $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 1$ or $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 0$. One consequence of defining the falsity conditions for $O^{\mathcal{P}}\phi$ in this way is that one cannot assert that the truth-teller is *not* paradoxical. But there is a way of strengthening the falsity conditions so as to accomplish this. One can say that $v^{<\alpha,\beta+1>}(O^{\mathcal{P}}\phi) = 0$ if $\forall v^{max} \in \mathcal{V}^{<\alpha,\beta>} : v^{max}(\phi) \neq \frac{1}{2}$, where v^{max} stands for a maximal KK-interpretation (cf. [10] for an equivalent definition). However, to avoid complications, we will stick to our official definition.

[31] It is worth highlighting that although every fixed-point for $O^{\mathcal{P}}$ is also a Kripke fixed-point, the converse doesn´t hold. There are Kripke fixed-points (fixed-points for $Tr(x)$) in $\mathcal{V}$ that are *not* fixed-points for $O^{\mathcal{P}}$, but these are needed to consistently interpret the operator. The same can be said about $O^{\mathcal{H}}$. For more details, see [27].

Successor ordinals for $O^{\mathcal{H}}$: $v^{<\alpha+1,0>}$ and $\mathcal{V}^{<\alpha+1,0>}$.

For every α, the micro-construction reaches a fixed-point, $v^{<\alpha,\beta>} = v^{<\alpha,\beta+1>}$ (this will be proved below). That fixed-point is used to define a new interpretation, $v^{<\alpha+1,0>}$, for sentences of the form $O^{\mathcal{H}}\phi$. So, for each ordinal α, let $v^{<\alpha+1,0>}$ be the interpretation obtained by letting sentences of the form $O^{\mathcal{H}}\phi$ behave as specified below and then applying the KK-operations:

$$v^{<\alpha+1,0>}(O^{\mathcal{H}}\phi) = \begin{cases} 1 & \text{if } \phi \in \mathcal{L}^{Tr} \text{ and } \exists v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 1 \text{ and } \exists v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 0 \\ 0 & \text{if } \forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = \text{i, for i} \in \{1, 0, \frac{1}{2}\}^{32} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

$$\mathcal{V}^{<\alpha+1,0>} = \{v : v \in \mathcal{V}^{<\alpha,\beta>} \text{ and } v \text{ is a KK-interpretation extending } v^{<\alpha+1,0>}\}$$

Limit ordinals for $O^{\mathcal{H}}$: $v^{<\lambda,0>}$ and $\mathcal{V}^{<\lambda,0>}$.

Let β be any ordinal. Then we let $v^{<\lambda,0>}$ be the interpretation obtained by letting sentences of the form $O^{\mathcal{H}}\phi$ behave as specified below and then applying the KK-operations:

$$v^{<\lambda,0>}(O^{\mathcal{H}}\phi) = \begin{cases} 1 & \text{if } \phi \in \mathcal{L}^{Tr} \text{ and } \exists v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \alpha < \lambda\} : v(\phi) = 1 \text{ and } \exists v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \alpha < \lambda\} : v(\phi) = 0 \\ 0 & \text{if } \forall v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \alpha < \lambda\} : v(\phi) = \text{i, for i} \in \{1, 0, \frac{1}{2}\} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

$$\mathcal{V}^{<\lambda,0>} = \{v : v \in \bigcap\{\mathcal{V}^{<\alpha,\beta>} : \alpha < \lambda\} \text{ and } v \text{ is a KK-interpretation extending } v^{<\lambda,0>}\}$$

If a KK-interpretation v satisfies the clause above, we will say that v is a *fixed-point for $O^{\mathcal{H}}$*.

Before moving on, we would like to explain in more detail why we needed to add in the truth-conditions for $O^{\mathcal{H}}\phi$ the extra requirement that $\phi \in \mathcal{L}^{Tr}$. Consider the sentences $O^{\mathcal{H}}\tau$ and $O^{\mathcal{H}}O^{\mathcal{H}}\tau$. In the first, the hypodoxicality operator applies to a sentence that does not contain any occurrences of itself (or of $O^{\mathcal{P}}$). In the second, however, it applies to a sentence that does contain such an occurrence. If we were to use the obvious characterization of hypodoxicality (i.e. the characterization without the extra requirement), it would be impossible to treat these two sentences differently at $v^{<1,0>}$. When we evaluate them at $v^{<1,0>}$, we need to look at the set of

[32] This should be understood as: $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 1$ or $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = 0$ or $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\phi) = \frac{1}{2}$.

KK-interpretations extending the fixed-point, $v^{<0,\gamma>}$, of the first micro-construction, that is, the set $\mathcal{V}^{<0,\gamma>}$. Yet, in $\mathcal{V}^{<0,\gamma>}$ there will be some KK-interpretations where τ is true and some KK-interpretations where τ is false. For $O^{\mathcal{H}}\tau$ the same holds—there will be some KK-interpretations where it is true and some where it is false. In other words, their behaviour across the set of KK-interpretations extending $v^{<0,\gamma>}$ is the same. As a result, the sentences $O^{\mathcal{H}}\tau$ and $O^{\mathcal{H}}O^{\mathcal{H}}\tau$ would receive the same truth-value at $v^{<1,0>}$ (they would both be true) and at every other stage of the construction. This suggests that the construction wouldn't be adequate—it would misdiagnose some sentences.

A second and more serious problem has to do with certain self-referential sentences involving $O^{\mathcal{H}}$. Consider the sentence that says of itself that it is hypodoxical—call it $\iota := O^{\mathcal{H}}Tr\ulcorner\iota\urcorner$. Let's assume that we're using the definition of hypodoxicality without the extra requirement. At $v^{<0,\gamma>}$ the sentence saying that ι is hypodoxical is neither true nor false. At $v^{<1,0>}$, however, the sentence will be true, because in $\mathcal{V}^{<0,\gamma>}$ there are some KK-interpretations where ι is true and some KK-interpretations where it is false. But since ι is precisely the sentence expressing that ι is hypodoxical, it follows that every KK-interpretations extending $v^{<1,0>}$ makes ι true. Thus, at $v^{<2,0>}$ ι will be false. So ι is unstable in a way that is at odds with the idea underpinning Kripke's original construction. In that setup, once a sentence is true (false) at a certain interpretation, it stays true (false) at all subsequent stages. Here, however, ι passes from truth to falsity, and, hence, the construction fails to be monotonic.

To overcome these difficulties what we've done is to define the truth-conditions for $O^{\mathcal{H}}$ in a way that $O^{\mathcal{H}}\phi$ is true only if ϕ belongs to the language $\mathcal{L}^{Tr}$, i.e. only sentences of $\mathcal{L}^{Tr}$ can be hypodoxical. This allows us to separate sentences like τ (i.e., sentences not containing occurrences of $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$) from sentences like $O^{\mathcal{H}}\tau$ and ι. τ will be hypodoxical, whereas $O^{\mathcal{H}}\tau$ and ι will not.[33] Of course, the resulting definition of $O^{\mathcal{H}}$ has a typed flavor.[34] This contrasts with our treatment of paradoxicality, which imposes no restrictions. There are sentences ϕ containing occurrences of the new operators such that $O^{\mathcal{P}}\phi$ is true.

Now, if our goal were merely to offer an account of paradoxicality and hypodoxicality that extensionally coincides with Kripke's treatment of such sentences, a typed approach to both operators would be more suitable. However, we believe that there is no a priori reason to think that these concepts should be typed. On the face of it, the notions of paradoxicality and hypodoxicality ought to be treated in roughly the same way that truth is.[35] Yet, in our framework there is something special about hypodoxicality that clashes with a completely unrestricted reading.

---

[33] We will meet the sentence ι again in Section 3.1.

[34] A consequence of this is that the account will have some expressive limitations. Consider the sentence $O^{\mathcal{H}}(\tau \wedge O^{\mathcal{H}}\tau)$. Intuitively, one would expect this sentence to be true, because one would expect $\tau \wedge O^{\mathcal{H}}\tau$ to behave hypodoxically. But since this sentence contains an occurrence of $O^{\mathcal{H}}$, $O^{\mathcal{H}}(\tau \wedge O^{\mathcal{H}}\tau)$ will not be true. However, all limitations of this kind pertain to sentences containing occurrences of the new operators. The account completely agrees with Kripke's theory for the fragment of the language without these operators.

[35] For a contrasting view, see [19, p. 714]. For discussion, see [27].

The general point is this. Intuitively, a sentence is hypodoxical if it can "go either way", i.e. it can consistently be interpreted as true or as false. Our construction is such that every sentence containing an occurrence of the new operators is like that at the start. So one needs some sort of restriction to avoid classifying every sentence as hypodoxical. To deal with this problem we've decided to impose a (very strong) condition in the definition of $O^{\mathcal{H}}\phi$. However, we are not committing ourselves to a fully typed framework. Note that the restriction only applies to the *truth*-conditions of $O^{\mathcal{H}}\phi$, not to its falsity conditions. There are many sentences ϕ containing occurrences of the new operators such that $O^{\mathcal{H}}\phi$ is *false* (some examples are offered in the next section). Ideally, a more symmetric (and less restrictive) treatment would be preferable, but at this point we are not sure if such a treatment can be offered.[36]

2.4 Some examples

The construction $\mathfrak{C}$ might seem a bit cumbersome at first sight. To a certain extent, we think that this is unavoidable. The introduction of the two operators—especially, $O^{\mathcal{H}}$—complicates things quite a bit, as we will shortly explain. In order to understand some of the features of $\mathfrak{C}$, it will be useful to go through a number of examples in some detail. Hopefully, this will serve to illustrate how $\mathfrak{C}$ works and also to show that it meets our expectations concerning $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$ in a large number of cases.

Our starting motivation was to develop a formal theory capable of diagnosing the liar, λ, and the truth-teller, τ, differently. So we will begin by taking a look at those sentences.

Example 1 *(The liar)*
Since $v^{<0,0>}$ is Kripke's minimal fixed-point, for every $v \in \mathcal{V}^{<0,0>} : v(\lambda) = \frac{1}{2}$. Therefore, at $v^{<0,1>}$ the truth-value of the sentence $O^{\mathcal{P}}\lambda$ will be 1, and it will remain to be so at every later stage.

Example 2 *(The truth-teller)*
The treatment of τ is different, however. There are some KK-interpretations extending $v^{<0,0>}$ where τ is true and some KK-interpretations extending $v^{<0,0>}$ where τ is false. But for every β, $v^{<0,\beta>}(O^{\mathcal{H}}\tau) = \frac{1}{2}$, since it is only at the first stage of the macro-construction that this sentence obtains a truth-value different from $\frac{1}{2}$. If $v^{<0,\gamma>}$ is a fixed-point of the first micro-construction, then there will be some KK-interpretations in $\mathcal{V}^{<0,\gamma>}$ where τ is true and some KK-interpretations in $\mathcal{V}^{<0,\gamma>}$ where τ is false. Given that $\tau \in \mathcal{L}^{Tr}$, it follows that at $v^{<1,0>}$ the truth-value of the sentence $O^{\mathcal{H}}\tau$ is 1, and it will remain that way at every later stage.

---

[36] Sometimes one cannot make progress if one waits for all related problems to be solved before proceeding. So here we are offering an admittedly imperfect solution to a difficult problem in the hope that a more satisfactory treatment will be available in future research. In connection to this, see Section 3.4.

Examples 1 and 2 show that it is not only the case that λ and τ are different as seen from the meta-language, but, crucially, this difference can be reflected in the object language by means of the sentence-classifying operators $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$. Moreover, $\mathfrak{C}$ is even more informative than that, as it also yields that $v^{<\alpha,\beta>}(O^{\mathcal{P}}\tau) \neq 1$ and that $v^{<\alpha,\beta>}(O^{\mathcal{H}}\lambda) \neq 1$, for every α and every β.

Sets of sentences can be adequately evaluated too. The sentences λ_1 and λ_2 generating the postcard paradox will be categorized as paradoxical, and the sentences λ_a and λ_b that compose the open pair will be evaluated as hypodoxical (see Footnote 1).

Example 3 *(The postcard paradox and the open pair)*
Since in every KK-interpretations v extending $v^{<0,0>}$, $v(\lambda_1) = v(\lambda_2) = \frac{1}{2}$, it follows that $v^{<0,1>}(O^{\mathcal{P}}\lambda_1) = v^{<0,1>}(O^{\mathcal{P}}\lambda_2) = 1$. A similar argument can be used to show that the sentences that conform the open pair are hypodoxical, i.e., $v^{<1,0>}(O^{\mathcal{H}}\lambda_a) = v^{<1,0>}(O^{\mathcal{H}}\lambda_b) = 1$. These assignments remain the same at later stages.

Other sets of sentences are handled in an analogous fashion. As for grounded sentences (in Kripke's sense), it is not hard to see that $\mathfrak{C}$ deems any attribution of paradoxicality or hypodoxicality to them false.

Example 4 *(Grounded sentences)*
If ϕ is a grounded sentence, then $v^{<0,0>}(\phi) \in \{1, 0\}$, since $v^{<0,0>}$ is Kripke's minimal fixed-point. This means that as we move through the sequence of interpretations, the value of ϕ remains the same. Thus, for every α and every $\beta > 0$, $v^{<\alpha,\beta>}(O^{\mathcal{P}}\phi) = 0$ and for every $\alpha > 0$ and every β, $v^{<\alpha,\beta>}(O^{\mathcal{H}}\phi) = 0$.

A family of more interesting cases pertains to sentences that contain iterations of $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$.

Example 5 *(Iterations I)*
Again, let ϕ be any grounded sentence. If $O^{\mathcal{P}} \dots O^{\mathcal{P}}\phi$ contains n iterations of $O^{\mathcal{P}}$, then at $v^{<0,n>}$ its value will be 0. And if $O^{\mathcal{H}} \dots O^{\mathcal{H}}\phi$ contains m iterations of $O^{\mathcal{H}}$, then at $v^{<m,0>}$ its value will be 0.

Iterations are also correctly explained for paradoxical and hypodoxical sentences. It is clear that although the liar is paradoxical, one would like to say that the sentence that expresses this claim, $O^{\mathcal{P}}\lambda$, is not itself paradoxical. Similarly, the truth-teller is hypodoxical, but one would like to assert that the sentence expressing this claim, $O^{\mathcal{H}}\tau$, is not itself hypodoxical. That is, one would expect $\mathfrak{C}$ to evaluate both $O^{\mathcal{P}}O^{\mathcal{P}}\lambda$ and $O^{\mathcal{H}}O^{\mathcal{H}}\tau$ as false.

Example 6 *(Iterations II)*
At $v^{<0,0>}$ and also at $v^{<0,1>}$ the truth-value of $O^{\mathcal{P}}O^{\mathcal{P}}\lambda$ is $\frac{1}{2}$, but at $v^{<0,2>}$ (and from there on) the sentence obtains the value 0, since at $v^{<0,1>}$ the sentence $O^{\mathcal{P}}\lambda$ obtains the value 1. More generally, a sentence of the form $O^{\mathcal{P}} \dots O^{\mathcal{P}}\lambda$ with n occurrences

of $O^{\mathcal{P}}$ will be $\frac{1}{2}$ until we reach $v^{<0,n>}$, where it acquires the value 0. A similar reasoning applies to $O^{\mathcal{H}}O^{\mathcal{H}}\tau$ and to more complicated sentences with iterations of $O^{\mathcal{H}}$.

The interaction of the two operators does not bring additional complications, as far as we can see. A couple of examples should be sufficient to assuage any potential concerns.

Example 7 *(Interactions)*
The sentence $O^{\mathcal{H}}O^{\mathcal{P}}\lambda$, e.g., receives the value $\frac{1}{2}$ at $v^{<0,\alpha>}$ for any α, but obtains the value 0 at $v^{<1,0>}$, since if $v^{<0,\gamma>}$ is a fixed-point of the first micro-construction, the sentence $O^{\mathcal{P}}\lambda$ has the value 1 at every interpretation extending $v^{<0,\gamma>}$. For another example, take the sentence $O^{\mathcal{P}}O^{\mathcal{P}}O^{\mathcal{H}}O^{\mathcal{H}}\lambda$. This sentence receives the value $\frac{1}{2}$ until one reaches the interpretation $v^{<2,2>}$, where it obtains the value 0. Other mixed sentences are treated analogously.

2.5 Some results

Now that we've offered a number of examples to illustrate how $\mathfrak{C}$ works, it is time to consider what some of its properties are. We start by noticing that as we move through the stages of $\mathfrak{C}$, some interpretations are rule out. A bit more formally:

Lemma 1 *For any ordinals α, α', β and β', if $\alpha < \alpha'$, or $\alpha = \alpha'$ but $\beta < \beta'$, then $\mathcal{V}^{<\alpha',\beta'>} \subseteq \mathcal{V}^{<\alpha,\beta>}$. In other words, the sets $\mathcal{V}$ form a decreasing sequence.*

This holds in virtue of the definition of each of the $\mathcal{V}$s. For some interpretation to be a member of $\mathcal{V}^{<\alpha',\beta'>}$, it has to be a member of $\mathcal{V}^{<\alpha,\beta>}$, the set corresponding to a prior stage.

One thing that could in principle occur in the present framework is for the set of $\mathcal{V}$s to get empty in the transition from one stage to the next. Let's suppose that at some stage the value of some sentence ϕ is 1, but at the next stage, its value is 0. Every KK-interpretation at the former stage will assign ϕ the value 1. But then the set of KK-interpretations at the latter stage will be empty, because every KK-interpretation at that stage should assign ϕ the value 0, but at the same time should be a subset of the set of KK-interpretations at the former stage, which is impossible.

In order to make sure that this doesn't happen in $\mathfrak{C}$, we need to show that once a sentence has obtained the value 1 or the value 0, it keeps that value at later stages. In other words, we need to show that $\mathfrak{C}$ is monotonic, in the following sense:

Lemma 2 *If $\alpha < \alpha'$, or $\alpha = \alpha'$ but $\beta < \beta'$, then for every sentence ϕ,*

- *If $v^{<\alpha,\beta>}(\phi) = 1$, then $v^{<\alpha',\beta'>}(\phi) = 1$, and*
- *if $v^{<\alpha,\beta>}(\phi) = 0$, then $v^{<\alpha',\beta'>}(\phi) = 0$,*

The proof of this result can be found in the appendix. In addition to guaranteeing that at no stage the set of $\mathcal{V}$s is empty, the monotonicity of $\mathfrak{C}$ entails the existence of a fixed-point of $\mathfrak{C}$, i.e. a fixed-point for $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$.

Proposition 1 *There is an interpretation, $v^{<\alpha,\beta>}$, such that for every subsequent interpretation, $v^{<\alpha',\beta'>}$, it holds that for every sentence ϕ, $v^{<\alpha,\beta>}(\phi) = v^{<\alpha',\beta'>}(\phi)$. That is, $v^{<\alpha,\beta>}$ is a fixed-point of $\mathfrak{C}$.*[37]

Note that if $v^{<\alpha,\beta>}$ is a fixed point of $\mathfrak{C}$, then the set of KK-interpretations extending $v^{<\alpha,\beta>}$ must be the same set as the set of KK-interpretations extending $v^{<\alpha',\beta'>}$. Hence, this is a fixed-point both for $O^{\mathcal{P}}$ and for $O^{\mathcal{H}}$ in that the set of KK-interpretations that must be considered in determining the truth-value of sentences of the form $O^{\mathcal{P}}\phi$ or of the form $O^{\mathcal{H}}\phi$ at a fixed-point of $\mathfrak{C}$ is the set of KK-interpretations extending that same fixed-point.

2.6 Meeting the desiderata

In what follows let's use the label v^{FP} for the KK-interpretation $v^{<\alpha,\beta>}$ that is a fixed-point for the construction $\mathfrak{C}$ in the sense discussed above. It is natural to define validity as preservation of truth in this interpretation.[38] Now, $\mathfrak{C}$ is always relativized to an interpretation v of the base language, $\mathcal{L}$. In this sense, there are as many possible starting points for $\mathfrak{C}$ as there are minimal fixed-points obtainable from interpretations for $\mathcal{L}$, all of which must be taken into account in the following definition:

Definition 4 (*Validity*) An argument from Γ to ϕ is valid ($\Gamma \models \phi$) if, and only if, for every interpretation v, the value of ϕ is 1 at v^{FP} whenever the value of each member of Γ is 1 at v^{FP}. A sentence ϕ is valid ($\models \phi$) if, and only if, for every interpretation v, the value of ϕ is 1 at v^{FP}.

Once this notion of validity is in place, we can show that $\mathfrak{C}$ meets the three desiderata that we laid down back in Section 2.3.

First, it is not hard to see that, according to $\mathfrak{C}$, the liar is paradoxical and the truth-teller is hypodoxical: for every v, $v^{FP}(O^{\mathcal{P}}\lambda) = v^{FP}(O^{\mathcal{H}}\tau) = 1$. In other words, we have both $\models O^{\mathcal{P}}\lambda$ and $\models O^{\mathcal{H}}\tau$. Also, paradoxicality and hypodoxicality behave differently in $\mathfrak{C}$: the liar is not hypodoxical and the truth-teller is not paradoxical. That is, $v^{FP}(O^{\mathcal{H}}\lambda) \neq 1$ and $v^{FP}(O^{\mathcal{P}}\tau) \neq 1$, which means that $\not\models O^{\mathcal{P}}\tau$ and $\not\models O^{\mathcal{H}}\lambda$.

Secondly, for any sentence ϕ not containing occurrences of $O^{\mathcal{P}}$ or $O^{\mathcal{H}}$, $\mathfrak{C}$ completely agrees with Kripke's account. That is, $\models O^{\mathcal{P}}\phi$ if and only if for every

[37] Since we are employing a quantifier-free language, the closure ordinal of $\mathfrak{C}$ is $\omega \times \omega$, and so the fixed-point of $\mathfrak{C}$ is $v^{<\omega,0>}$. But for richer languages the closure ordinal could be different. See also Footnote 27.

[38] Natural, yes, but of course there are other options in the table. For a brief exploration of some of these options, see [27].

interpretation extending v^{FP}, the value of ϕ at that interpretation is $\frac{1}{2}$, which means that ϕ is paradoxical in Kripke's sense. Also, $\models O^{\mathcal{H}}\phi$ if and only if there is an interpretation extending v^{FP} at which ϕ is true and there is another interpretation extending v^{FP} at which ϕ is false, which means that ϕ is hypodoxical in the sense of Section 2.1.

Thirdly, we can show that the new operators interact in the following way with the logical connectives (and with themselves):

Proposition 2 *Some facts about $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$:*

1. $O^{\mathcal{P}}\phi \models O^{\mathcal{P}}\neg\phi$, *but* $\neg O^{\mathcal{P}}\phi \not\models O^{\mathcal{P}}\neg\phi$.
2. $O^{\mathcal{P}}\neg\phi \models O^{\mathcal{P}}\phi$, *but* $O^{\mathcal{P}}\neg\phi \not\models \neg O^{\mathcal{P}}\phi$.
3. $O^{\mathcal{H}}\phi \models O^{\mathcal{H}}\neg\phi$, *but* $\neg O^{\mathcal{H}}\phi \not\models O^{\mathcal{H}}\neg\phi$.
4. $O^{\mathcal{H}}\neg\phi \models O^{\mathcal{H}}\phi$, *but* $O^{\mathcal{H}}\neg\phi \not\models \neg O^{\mathcal{H}}\phi$.
5. $O^{\mathcal{P}}\phi \wedge O^{\mathcal{P}}\psi \models O^{\mathcal{P}}(\phi \wedge \psi)$, *but* $O^{\mathcal{P}}(\phi \wedge \psi) \not\models O^{\mathcal{P}}\phi \wedge O^{\mathcal{P}}\psi$.
6. $O^{\mathcal{H}}\phi \wedge O^{\mathcal{H}}\psi \not\models O^{\mathcal{H}}(\phi \wedge \psi)$ *and* $O^{\mathcal{H}}(\phi \wedge \psi) \not\models O^{\mathcal{H}}\phi \wedge O^{\mathcal{H}}\psi$.
7. $O^{\mathcal{P}}(\phi \vee \psi) \not\models O^{\mathcal{P}}\phi \vee O^{\mathcal{P}}\psi$ *and* $O^{\mathcal{P}}\phi \vee O^{\mathcal{P}}\psi \not\models O^{\mathcal{P}}(\phi \vee \psi)$.[39]
8. $O^{\mathcal{H}}(\phi \vee \psi) \models O^{\mathcal{H}}\phi \vee O^{\mathcal{H}}\psi$, *but* $O^{\mathcal{H}}\phi \vee O^{\mathcal{H}}\psi \not\models O^{\mathcal{H}}(\phi \vee \psi)$.
9. $O^{\mathcal{P}}O^{\mathcal{P}}\phi \models O^{\mathcal{P}}\phi$, *but* $O^{\mathcal{P}}\phi \not\models O^{\mathcal{P}}O^{\mathcal{P}}\phi$.
10. $O^{\mathcal{H}}O^{\mathcal{H}}\phi \models O^{\mathcal{H}}\phi$, *but* $O^{\mathcal{H}}\phi \not\models O^{\mathcal{H}}O^{\mathcal{H}}\phi$.

There are a lot of facts about paradoxicality and hypodoxicality up there, and it would be extremely tedious to go through all of them in detail. But let us briefly discuss some of them.

We take it that, in this setting, 1-4 are indisputable facts about paradoxicality and hypodoxicality. For example, it is easy to see that $O^{\mathcal{P}}\phi \models O^{\mathcal{P}}\neg\phi$ holds. If $v^{FP}(O^{\mathcal{P}}\phi) = 1$, then $\forall v \in \mathcal{V}^{FP} : v(\phi) = \frac{1}{2}$. Hence, $\forall v \in \mathcal{V}^{FP} : v(\neg\phi) = \frac{1}{2}$, since every v is a KK-interpretation. From this it follows that $v^{FP}(O^{\mathcal{P}}\neg\phi) = 1$.

That the first item of 5 holds can be shown as follows. Assume that $v^{FP}(O^{\mathcal{P}}\phi) = 1$ and $v^{FP}(O^{\mathcal{P}}\psi) = 1$. This holds if and only if $\forall v \in \mathcal{V}^{FP} : v(\phi) = \frac{1}{2}$ and $\forall v \in \mathcal{V}^{FP} : v(\psi) = \frac{1}{2}$. Since all vs are KK-interpretations, the previous claims hold only if $\forall v \in \mathcal{V}^{FP} : v(\phi \wedge \psi) = \frac{1}{2}$. And, finally, that is the case if and only if $v^{FP}(O^{\mathcal{P}}(\phi \wedge \psi)) = 1$. The second item of 5 fails if, for instance, ϕ is λ, and ψ is $\top$ (or any other sentence that is true at every KK-interpretation).

In the case of 6, neither implication holds. A counterexample to the first is $O^{\mathcal{H}}\tau \wedge O^{\mathcal{H}}\neg\tau \not\models O^{\mathcal{H}}(\tau \wedge \neg\tau)$ and a counterexample to the second is $O^{\mathcal{H}}(\tau \wedge \top) \not\models O^{\mathcal{H}}\tau \wedge O^{\mathcal{H}}\top$.

The first item in 7 is the least obvious. A counterexample to it can be constructed by letting ϕ be $(\tau\wedge\lambda)\wedge\neg(\tau\wedge\lambda)$ and letting ψ be $(\neg\tau\wedge\lambda)\wedge\neg(\neg\tau\wedge\lambda)$. One thus obtains

[39] Disjunction is not part of the official language, but it can be defined in terms of conjunction in the usual way: $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$. Thus, these facts should be translated appropriately.

the awful-looking $O^{\mathcal{P}}(((\tau\wedge\lambda)\wedge\neg(\tau\wedge\lambda))\vee((\neg\tau\wedge\lambda)\wedge\neg(\neg\tau\wedge\lambda)))\not\models O^{\mathcal{P}}((\tau\wedge\lambda)\wedge\neg(\tau\wedge\lambda))\vee O^{\mathcal{P}}((\neg\tau\wedge\lambda)\wedge\neg(\neg\tau\wedge\lambda))$. In this case, since $\forall v\in\mathcal{V}^{FP}: v(\phi\vee\psi)=\frac{1}{2}$, $v^{FP}(O^{\mathcal{P}}(\phi\vee\psi))=1$, but it is not the case that $\forall v\in\mathcal{V}^{FP}: v(\phi)=\frac{1}{2}$ nor is it the case that $\forall v\in\mathcal{V}^{FP}: v(\psi)=\frac{1}{2}$.[40] A counterexample to its converse (7's second item) is $O^{\mathcal{P}}\lambda\vee O^{\mathcal{P}}\top\not\models O^{\mathcal{P}}(\lambda\vee\top)$.

As for 8, its first item holds. We reason contrapositively. If $v^{FP}(O^{\mathcal{H}}\phi\vee O^{\mathcal{H}}\psi)=0$, then $\forall v\in\mathcal{V}^{FP}: v(\phi)=i$, for $i\in\{1,0,\frac{1}{2}\}$ and $\forall v\in\mathcal{V}^{FP}: v(\psi)=i$, for $i\in\{1,0,\frac{1}{2}\}$. It follows that $\forall v\in\mathcal{V}^{FP}: v(\phi\vee\psi)=i$, for $i\in\{1,0,\frac{1}{2}\}$. Hence, $v^{FP}(O^{\mathcal{H}}(\phi\vee\psi))=0$ too. If, instead, $v^{FP}(O^{\mathcal{H}}\phi\vee O^{\mathcal{H}}\psi)=\frac{1}{2}$, then neither ϕ nor ψ are hypodoxical, and they cannot both be grounded. It is not hard to see that this implies that either $\exists v\in\mathcal{V}^{FP}: v(\phi\vee\psi)=\frac{1}{2}$, or $\forall v\in\mathcal{V}^{FP}: v(\phi\vee\psi)=i$ for $i\in\{1,0,\frac{1}{2}\}$. And this means that $v^{FP}(O^{\mathcal{H}}(\phi\vee\psi))\neq 1$. A counterexample to the second item in 8 is $O^{\mathcal{H}}\tau\vee O^{\mathcal{H}}\top\not\models O^{\mathcal{H}}(\tau\vee\top)$.

Items 9 and 10 deal with iterations of our operators. For instance, the liar is paradoxical, so it holds that $O^{\mathcal{P}}\lambda$, but this sentence is not itself paradoxical, so it is not the case that $O^{\mathcal{P}}O^{\mathcal{P}}\lambda$. Consequently, $O^{\mathcal{P}}\phi\not\models O^{\mathcal{P}}O^{\mathcal{P}}\phi$. Similarly, the truth-teller is hypodoxical, but the sentence expressing this fact is not itself hypodoxical, thus $O^{\mathcal{H}}\phi\not\models O^{\mathcal{H}}O^{\mathcal{H}}\phi$. The two converse facts hold. No sentence of the form $O^{\mathcal{P}}O^{\mathcal{P}}\phi$ can be 1 at v^{FP}. A fortiori, $O^{\mathcal{P}}O^{\mathcal{P}}\phi\models O^{\mathcal{P}}\phi$, for no interpretation makes the premise true. Similarly, $O^{\mathcal{H}}O^{\mathcal{H}}\phi\models O^{\mathcal{H}}\phi$, since no sentence of the form $O^{\mathcal{H}}O^{\mathcal{H}}\phi$ can be 1 at v^{FP}.

We've listed a number of principles and non-principles of $\mathfrak{C}$, but there are many other facts about paradoxicality and hypodoxicality that we haven't considered. Surely, a more systematic approach would be preferable. In particular, it would be interesting to see what would be required for a proof-theoretic treatment of the operators $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$. But it is highly non-trivial how to do this, so we must leave the matter for another day.[41] We hope though that the facts above are sufficient to give us some insights into how one should understand paradoxicality and hypodoxicality, and to show that the three desiderata that we've proposed for theories of these concepts can be met to a certain extent in our account.

3 Objections and open problems

3.1 Self-referential sentences

We've pointed out that the introduction of the operators $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$ brings with it the possibility of expressing new self-referential sentences, yet we haven't so far

[40] This example—involving a paradoxical disjunction with non-paradoxical disjuncts—is interesting for a different but related reason, cf. [23].

[41] Still, see [20] for some potential problems affecting a proof-theoretic treatment of paradoxicality; and see [26] and [27] for replies.

discussed how $\mathfrak{C}$ diagnoses these sentences. For example, there will be a sentence, π, saying of itself that it is paradoxical, and a sentence, ι, saying of itself that it is hypodoxical, among many others:

$$\pi := O^{\mathcal{P}} Tr\ulcorner \pi \urcorner$$
$$\iota := O^{\mathcal{H}} Tr\ulcorner \iota \urcorner$$

As it stands, $\mathfrak{C}$ deems both sentences neither true nor false, but one could find this unconvincing. It might be somewhat natural or even intuitive—assuming that we have intuitions about such things—to think that claims about paradoxicality and hypodoxicality ought to behave bivalently. According to Kripke's theory, a sentence is either paradoxical or it isn't, and a sentence is either hypodoxical or it isn't. Moreover, one might suggest that ι is true on the grounds that it can take any truth-value (in some appropriate sense of 'can') and that that's precisely what it says. Also, it is not implausible to think that π is false on the grounds that there's nothing paradoxical about a sentence saying of itself that it is paradoxical. We won't try to analyze whether these diagnosis are plausible or not, and we won't take a stand on which treatment of these (and similar) sentences is preferable.[42] Yet, we will mention that there is a straightforward way of amending the construction $\mathfrak{C}$ so as to evaluate these sentences in a classical way. More specifically, what we can do is to 'close-off' $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$, and thereby transform them into bivalent operators, so that any attribution of paradoxicality or hypodoxicality will be either true or false.[43]

One way of partially 'classicizing' v^{FP} consists in defining a new interpretation, call it v^{BIV}, that can be seen as a kind of classical close-off of v^{FP} for paradoxicality and hypodoxicality. The rough idea is to say, first, that a sentence $O^{\mathcal{P}}\phi$ is true at v^{BIV} if $\forall v \in \mathcal{V}^{FP} : v(\phi) = \frac{1}{2}$, and *is false otherwise*. Secondly, we can say that a sentence $O^{\mathcal{H}}\phi$ is true at v^{BIV} if $\exists v \in \mathcal{V}^{FP} : v(\phi) = 1$ and $\exists v \in \mathcal{V}^{FP} : v(\phi) = 0$, and *is false otherwise*. Clearly, π will be false at v^{BIV} and ι will be true.

But there are costs attached to the classicization. For one thing, it is not longer the case that the value of sentences of the form $O^{\mathcal{P}}\phi$ or $O^{\mathcal{H}}\phi$ at v^{BIV} will depend on the behaviour of ϕ at the set of interpretations extending v^{BIV}. As a consequence, v^{BIV} is no longer a fixed-point for them. For another, v^{BIV} classifies a number of sentences as paradoxical or hypodoxical when they are arguably not so. For example, to the extent that there are reasons for believing that ι is true, we think that these reasons should also support the claim that the following sentence is false: $O^{\mathcal{H}}\neg O^{\mathcal{H}}\iota$. For if $O^{\mathcal{H}}\iota$ is true, its negation, $\neg O^{\mathcal{H}}\iota$, is false. So one would expect $O^{\mathcal{H}}\neg O^{\mathcal{H}}\iota$ to be false. Yet, this sentence is true at v^{BIV}.

At any rate, it is not our goal to argue in favor of v^{FP} over v^{BIV} or vice versa. All we're hoping to accomplish is to clearly lay out the two options with some of their costs and benefits.

---

[42] We do wish to note, though, that it is rather troublesome to rely on intuitions in the assessment of π, ι and related sentences. These sentences are about two highly theoretical concepts, paradoxicality and hypodoxicality, so it could be argued that there are in fact no intuitions about them.

[43] The idea of 'closing-off' $Tr(x)$ is due to Kripke [19, p. 715]. Of course, here we are only closing-off $O^{\mathcal{P}}$ and $O^{\mathcal{H}}$, not $Tr(x)$, so the resulting language is not completely classical, as in Kripke's case.

3.2 Contingent paradoxes (and contingent hypodoxes)

One matter that we haven't discussed at all so far is how to deal with contingent paradoxes. The idea of a contingent paradox is familiar. Informally speaking, a sentence is usually said to be contigently paradoxical if its paradoxicality depends on contingent matters. A typical example of this is a sentence $\lambda \vee p$, which is paradoxical contingent on p not being true. If p is true, then $\lambda \vee p$ is simply true and, hence, not paradoxical.[44]

We think that contingent paradoxes do not pose a problem for the present account. In fact, the account arguably yields the right diagnosis for these sentences. The value of the sentence $O^{\mathcal{P}}(\lambda \vee p)$ will depend on the value of p in the expected way. There are interpretations at which p is true and interpretations at which it is false. If v is an interpretation at which p is false, then v^{FP} will make $O^{\mathcal{P}}(\lambda \vee p)$ true, that is, it will declare $\lambda \vee p$ paradoxical. If, instead, v is an interpretation at which p is true, then v^{FP} will make $O^{\mathcal{P}}(\lambda \vee p)$ false, i.e. it will declare $\lambda \vee p$ unparadoxical. We can thus say that a sentence ϕ is *contingently paradoxical* if there are some interpretations v such that $v^{FP}(O^{\mathcal{P}}\phi) = 1$ and there are some interpretations v such that $v^{FP}(O^{\mathcal{P}}\phi) \neq 1$.

Kripke famously pointed out [19, p. 691] that "many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features". We submit that most of our ordinary assertions about truth and falsity are also liable to exhibit *hypodoxical* features. The idea of a contingent hypodox is of course less familiar. But, informally speaking, we can say that a sentence is contingently hypodoxical if its hypodoxicality depends on contingent matters. In other words, a sentence may exhibit hypodoxical behaviour—i.e. there's no non-arbitrary way of evaluating it as true or as false—if the empirical facts are in a certain way. For example, the sentence $\tau \vee p$ is hypodoxical contingent on p being false. If p is true, then $\tau \vee p$ is simply true and, hence, not hypodoxical. We can thus say, a bit less informally, that a sentence ϕ is *contingently hypodoxical* if there are some interpretations v such that $v^{FP}(O^{\mathcal{H}}\phi) = 1$ and there are some interpretations v such that $v^{FP}(O^{\mathcal{H}}\phi) \neq 1$. If one starts with an interpretation v at which p is false, then $v^{FP}(O^{\mathcal{H}}(\tau \vee p)) = 1$. If, however, one starts with an interpretation v at which p is true, then $v^{FP}(O^{\mathcal{H}}(\tau \vee p)) = 0$. Thus, the sentence $\lambda \vee p$ is contingently hypodoxical. In view of this and similar examples, we can say that the theory we've proposed has the resources to diagnose contingent paradoxes and contingent hypodoxes correctly.

---

[44] It could be suggested that a sentence such as $\lambda \vee p$ is paradoxical regardless of whether p is true or not. If, for instance, one doesn't know whether p is the case, then one may say that $\lambda \vee p$ is paradoxical because it is possible to use this sentence to infer p regardless of what p says. The idea underpinning this suggestion is that paradoxicality is ultimately an epistemic notion. A sentence may be both true and paradoxical-relative-to-a-speaker. Here we are not understanding paradoxicality in that way, but rather as a property that some sentences possess purely in virtue of how they behave across a certain set of interpretations. It would be interesting to explore what are the connections between these two ways of understanding paradoxicality, but we must leave a discussion of this and related matters for another occasion.

3.3 Extensionality and other ungrounded sentences

One might have the impression that the complicated formal machinery we've deployed to distinguish paradoxicality from hypodoxicality can be dispensed with if instead of employing three-valued interpretations, one were to use an FDE-based theory, with four-valued interpretations (cf., e.g., Albert Visser's [30]). Since now one has 'true', 'false', 'true and false', and 'neither true nor false' as semantic categories, one can separate liars from truth-tellers extensionally (i.e. within a single model). Moreover, in a four-valued framework it might even be possible to define paradoxicality and hypodoxicality operators such that $O^{\mathcal{P}}\phi$ is true if and only if ϕ is both true and false, and $O^{\mathcal{H}}\phi$ is true if and only if ϕ is neither true nor false.

However, we think that this approach faces several problems. First, arguably these operators only capture the notions of 'being true and false at an interpretation' and that of 'being neither true nor false at an interpretation', respectively. But, for example, the concept of paradoxicality, as Kripke defines it, requires something else entirely, due to its intensional character. To know whether a sentence $O^{\mathcal{P}}\phi$ is true at some interpretation it is not enough to know what the value of ϕ is at that interpretation. One needs to know what the value of ϕ is at other interpretations as well. We believe that this is true even in a four-valued setting.

Secondly, if one likes the idea that semantic properties supervene on non-linguistic facts, then a four-valued FDE-based framework might not be the best way to go. In fact, because of the way in which conjunction and disjunction behave in FDE, the resulting theory of truth appears to be at odds with the supervenience thesis. For example, $\lambda \vee \tau$ will be strictly true in virtue of the interaction of the values 'both true and false' and 'neither true nor false', but there is no non-linguistic fact on which its (strict) truth depends.[45]

Third, there are other types of sentences besides liars and truth-tellers that do not seem to straightforwardly fit any of the four semantic categories available in FDE. Theories of truth tend to focus almost exclusively on paradoxes, and thus they leave other ungrounded sentences as a curious and somewhat second-class phenomenon. We've gone some way towards remedying this situation by offering a formal treatment of the notion of hypodoxicality, but there are ungrounded sentences that are neither paradoxical nor hypodoxical. For example, consider the following two:

$$\alpha := Tr\ulcorner\alpha\urcorner \vee \neg Tr\ulcorner\alpha\urcorner \qquad\qquad \beta := Tr\ulcorner\beta\urcorner \wedge \neg Tr\ulcorner\beta\urcorner$$

The first one, α, asserts of itself that it is either true or untrue, whereas the second one, β, asserts of itself that it is both true and untrue.[46] These sentences are neither

[45] The argument is due to Luca Castaldo and we've borrowed it from his paper [5]. For a discussion on the semantic supervenience thesis in languages containing semantic predicates other than Tr, cf. [15].

[46] There is a notational ambiguity between the ordinals α and β, and the sentences α and β (as there was between the ordinal λ and the sentence λ), but we trust that the context is always sufficient to disambiguate appropriately.

grounded nor paradoxical. α is a sentence that cannot be false, whereas β is a sentence that cannot be true. They are, in the terminology of Cook and Tourville's [10], semi-true and semi-false (respectively).[47]

In view of these problems, we think that the four-valued approach shouldn't be accepted, and that an intensional proposal is preferable. But wait a minute! A moment's reflection shows that on account of the third problem, our proposal could be said to face an objection that is quite similar to the one that we have posed to standard truth-theoretic approaches and that has been the driving force behind this paper. We've stressed that these approaches fail to make a distinction between ungrounded sentences that are paradoxical and ungrounded sentences that are hypodoxical. Yet, now it seems that our account fails to make a distinction between semi-true sentences like α and semi-false sentences like β. In the construction $\mathfrak{C}$, $v^{FP}(O^{\mathcal{P}}(\alpha)) = v^{FP}(O^{\mathcal{P}}(\beta)) = \frac{1}{2}$ and $v^{FP}(O^{\mathcal{H}}(\alpha)) = v^{FP}(O^{\mathcal{H}}(\beta)) = \frac{1}{2}$, so the account is silent on the semantic status of these sentences. This has some bearing on the general project that we've mentioned back in Section 2—to offer an account that is fine-grained enough to distinguish different semantic pathologies. Even if we avail ourselves of the paradoxicality and the hypodoxicality operators, we won't be able to positively characterize sentences like these. Hence, the problem is that we have no way of (positively) talking about them without resorting to the meta-language.

However, unlike in the case of the FDE-based approach, where one would have to add new semantic categories to deal with these sentences, in the present framework we can get around this difficulty by the introduction of two primitive operators for semi-truth and semi-falsity, specifically designed to characterize sentences like α and β. We won't explain how to do this here because, fortunately for us, this has already been done by Cook and Tourville (cf. [10]). With these additional operators at our disposal, the objection can be met.[48]

---

[47] Cook and Tourville develop an account of paradoxicality, semi-truth, and semi-falsity—although not of hypodoxicality—as part of their embracing revenge view (cf. Footnote 5). The present account differs from their approach in that they only consider operators that are monotonic in a novel technical sense that they introduce. $O^{\mathcal{H}}$ turns out not to be monotonic in that way. Thus, $O^{\mathcal{H}}$ is not available on their approach and, as a consequence, the semantic status of τ and of sentences like it can only be explained negatively. That is, one can say, for example, that they are not paradoxical, but there is no operator or predicate that is true of hypodoxical sentences and only of them. It is possible to define in their approach a 'semi-classicality' operator that applies to truth-teller like sentences, but it picks out a property that also applies to grounded sentences and to other ungrounded sentences that are not hypodoxical, such as α and β.

[48] There are complications that we are deciding to ignore here. Cook and Tourville's operators fail to classify ungrounded sentences like $\lambda \vee \tau$ (and $\lambda \wedge \tau$) that are true (false) at certain fixed-points, neither true nor false at other fixed-points, and false (true) at no fixed-point. For a discussion of this issue, see [27].

3.4 Self-reference and arithmetic

Since we've restricted ourselves to a propositional framework, this might raise the suspicion that paradoxes are averted by making the underlying theory too weak—in the sense that the theory only manages to avoid triviality by being incapable of expressing some paradoxical sentences. In fact, a reasonable constraint on a theory of truth is that it ought to contain an account of the objects to which truth can be applied. For this reason, theories of truth are typically couched in languages that have the resources to develop their own syntax. Peano arithmetic and other sufficiently expressive theories are suitable for this task because they contain a syntactic theory for sentences, which are taken to be the bearers of truth, via a Gödel coding.

In Section 2.2 we employed a rather unusual way of constructing self-referential sentences. One reason for this is that we wanted to give an account of contingent paradoxes (and contingent hypodoxes), and in order to do this the base language must contain contingent sentences. But there is a technical and more substantial reason for not using an arithmetical language. We've seen that in the truth-conditions for $O^{\mathcal{H}}$ the idea is, roughly, that a sentence $O^{\mathcal{H}}\phi$ is true only if $\phi \in \mathcal{L}^{Tr}$. The point of this condition was to separate sentences like τ from sentences like $O^{\mathcal{H}}\tau$ or $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$. We wanted to say that the first one is hypodoxical, but that the other two are not. The way in which we managed to do this was by treating sentences such as τ differently from sentences like $O^{\mathcal{H}}\tau$ or $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$ on account of the fact that the former doesn't contain an (explicit or implicit) occurrence of $O^{\mathcal{H}}$, but the latter two do. Our way of providing names for the sentences of the language allowed us to separate these types of sentences in a rigorous way. The sentence τ belongs to $\mathcal{L}^{Tr}$, whereas the sentences $O^{\mathcal{H}}\tau$ and $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$ belong to $\mathcal{L}^{+}$ but not to $\mathcal{L}^{Tr}$.

If, instead, the language $\mathcal{L}^{+}_{PA}$ of Peano arithmetic (enriched with a truth predicate and operators for paradoxicality and hypodoxicality) is employed, then it is unclear to us whether the distinction can be made in a satisfactory way. The problem is that sentences may contain implicit occurrences of the new operators. In a sentence like $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$, the expression $\ulcorner O^{\mathcal{H}}\tau\urcorner$ is a singular term standing for a numeral. That is, $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$ is actually of the form $Tr(\overline{n})$ where $\overline{n}$ is the numeral that denotes the number n and n is the Gödel code of the sentence $O^{\mathcal{H}}\tau$ under some appropriate Gödel numbering. One could hope that there is always a way of recursively tracking down sentences that, like $Tr\ulcorner O^{\mathcal{H}}\tau\urcorner$, contain an implicit occurrence of $O^{\mathcal{H}}$ (or of $O^{\mathcal{P}}$). But if quantified sentences are available, it is not obvious whether this can be done satisfactorily. By way of example, it is possible to construct a sentence of $\mathcal{L}^{+}_{PA}$ that doesn't contain any explicit or implicit occurrences of $O^{\mathcal{H}}$ or of $O^{\mathcal{P}}$ but that is equivalent to a sentence that does contain an occurrence of one of these operators. Consider the sentence $\exists x(Tr(x) \wedge x = \ulcorner O^{\mathcal{H}}\tau\urcorner)$, which contains an implicit occurrence of $O^{\mathcal{H}}$. Now let $\phi(x)$ be a unary formula wherein (unlike in the case of the formula $x = \ulcorner O^{\mathcal{H}}\tau\urcorner$) the numeral $\ulcorner O^{\mathcal{H}}\tau\urcorner$ does not occur explicitly or implicitly but such that $\forall x(\phi(x) \to x = \ulcorner O^{\mathcal{H}}\tau\urcorner)$ is provable in Peano arithmetic. That is, $\phi(x)$ is true of $\ulcorner O^{\mathcal{H}}\tau\urcorner$ and only of this term. Then, clearly, the sentence $\exists x(Tr(x) \wedge \phi(x))$ must be equivalent to $\exists x(Tr(x) \wedge x = \ulcorner O^{\mathcal{H}}\tau\urcorner)$. The problem is that we now face a dilemma. On the one hand, if we say that the former does not contain an implicit occurrence of $O^{\mathcal{H}}$, our theory will diagnose these two sentences differently merely in virtue of the syntactic fact that one contains an (implicit) occurrence of $O^{\mathcal{H}}$. The first will come out as hypodoxical and the second one will not, in spite of their equivalence. On the other hand, if we say that $\exists x(Tr(x) \wedge \phi(x))$ contains an (implicit) occurrence of $O^{\mathcal{H}}$, perhaps on the grounds that it quantifies over a sentence containing an occurrence of it, then we need to offer a rigorous definition of implicit occurrence for $\mathcal{L}^{+}_{PA}$, a highly non-trivial feat. If such a definition can be given, then the framework should be adapted accordingly.[49] If such a definition cannot be given, then a more substantial modification of the framework is needed.

We believe that there are at least two different thoughts that might be worth considering in connection to this. First, one could transpose the entire construction into the context of an infinitary propositional language, where a definition of 'occurrence' can be given in a straightforward manner. We think that it shouldn't be too difficult to adapt the construction in this way. Second, one could develop an alternative version of $\mathfrak{C}$, which is a full-fledged revision semantics, rather than a fixed-point semantics. However, doing this properly is far from easy, and so we only mention the possibility here.

4 Concluding remarks

These and other questions remain open. However, our aim in this paper was modest: to provide a fine-grained treatment of truth that successfully reflects the distinction between paradoxical and hypodoxical sentences. In order to accomplish that, we have shown how to construct a framework that extends Kripke's fixed-point account with two primitive operators, one standing for paradoxicality and the other for hypodoxicality. What is of value in our approach, we reckon, is that it succeeds in characterizing the semantic status of many sentences without employing metalinguistic resources, and that it does so in a way that avoids conflating sentences that are paradoxical with sentences that are hypodoxical.

Acknowledgements We started working on this paper a few years ago. We submitted an earlier version of it for publication back in 2019 and a referee noted that there was a problem with the definition of hypodoxicality. We then put it aside for some time. Some of its contents found their way into another paper written by one of us (LR), "Paradoxicality without Paradox", which was then published. After some time we decided to fix the problem noted by the referee, since we thought that the paper was still a valuable addition to the literature. We know of no other attempt at characterizing the notion of hypodoxicality formally, and we are aware of only a handful of proposals that care about making the distinction between liars and truth-tellers in the object language. Earlier drafts of the paper were presented at conferences in Buenos Aires, Rosario and Campinas. We are grateful to the organizers and audiences of these conferences. We would especially like to thank Eduardo Barrio, Roy Cook, Bruno Da Ré, Leon Horsten, Mattia Petrolo, Damián Szmuc, Paula

---

[49] One possibility worth exploring is to offer a characterization of 'implicit occurrence' along the lines of Lavinia Picollo's account of alethic reference in e.g. [21]. However, pursuing this idea is beyond the scope of this paper.

Teijeiro, Giorgio Venturi, and several anonymous reviewers for very useful feedback. This work was supported by PLEXUS (grant agreement no 101086295), a Marie Sklodowska-Curie action funded by the EU under the Horizon Europe Research and Innovation Programme and also by Fundaçao para a Ciencia e a Tecnologia via project 2022.03194. PTDC, New perspectives on the objects and grounds of the structural rules.

Appendix

The monotonicity of $\mathfrak{C}$

In this appendix we prove that the interpretations of the construction $\mathfrak{C}$ are monotonically increasing.

Lemma 3 (Monotonicity) *If $\alpha < \alpha'$, or $\alpha = \alpha'$ and $\beta < \beta'$, then for every sentence ϕ,*

- *If $v^{<\alpha,\beta>}(\phi) = 1$, then $v^{<\alpha',\beta'>}(\phi) = 1$, and*
- *if $v^{<\alpha,\beta>}(\phi) = 0$, then $v^{<\alpha',\beta'>}(\phi) = 0$,*

Proof We only offer a sketch of the proof. We need to establish, first, that:

1. If $v^{<\alpha,\beta>}(\phi) = 1(0)$, then $v^{<\alpha,\beta+1>}(\phi) = 1(0)$.
2. If $\forall\beta < \lambda : v^{<\alpha,\beta>}(\phi) = 1(0)$, then $v^{<\alpha,\lambda>}(\phi) = 1(0)$.

Items 1 and 2 guarantee that for each ordinal α, the micro-construction starting from $v^{<\alpha,0>}$ is monotonic and thus reaches a fixed-point, $v^{<\alpha,\gamma>} = v^{<\alpha,\gamma+1>}$, for $O^{\mathcal{P}}$. Once this fact is in the bag, establishing the following two claims is enough to prove that the macro-construction is monotonic as well:

3. If $v^{<\alpha,\gamma>}(\phi) = 1(0)$, then $v^{<\alpha+1,0>}(\phi) = 1(0)$.
4. If $\forall\alpha < \lambda : v^{<\alpha,\beta>}(\phi) = 1(0)$, then $v^{<\lambda,0>}(\phi) = 1(0)$.

And claims 1-4 are sufficient to establish our main result. The proof, as usual, is by induction on the complexity of ϕ. We'll just look at successor ordinals (i.e. items 1 and 3). The reasoning for limit ordinals (items 2 and 4) is similar. We start with item 1: if $v^{<\alpha,\beta>}(\phi) = 1(0)$, then $v^{<\alpha,\beta+1>}(\phi) = 1(0)$.

- *ϕ is an atomic sentence*. Then it is either a propositional variable p or it is a truth predication $Tr(b)$. If it is the former, then its truth-value does not change across the sequence of interpretations. If it is a truth predication, then either it contains an implicit occurrence of one of the new operators or it doesn't. If it doesn't, then its truth-value does not change across the sequence of interpretations. If it does, since every v is a KK-interpretation (i.e., a fixed-point), the truth-value of $Tr(b)$ will ultimately depend on the truth-value of a sentence that is not a truth predication, so this case reduces to one of the other cases.
- *ϕ is one of $\neg\psi$, $\psi \wedge \chi$, $O^{\mathcal{H}}\psi$, or $O^{\mathcal{P}}\psi$.*
  - For the logical connectives we use the inductive hypothesis in the usual way. Just to illustrate, take the case of negation. If $v^{<\alpha,\beta>}(\neg\psi) = 1(0)$, then $v^{<\alpha,\beta>}(\psi) = 0(1)$. By the inductive hypothesis, $v^{<\alpha,\beta+1>}(\psi) = 0(1)$. So, $v^{<\alpha,\beta+1>}(\neg\psi) = 1(0)$.

– The only interesting cases are those involving $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$. But note that $O^{\mathcal{H}}$-sentences do not change their value in the transition from $v^{<\alpha,\beta>}$ to $v^{<\alpha,\beta+1>}$, so we only need to look at sentences of the form $O^{\mathcal{P}}\psi$. Let's assume, then, that $v^{<\alpha,\beta>}(O^{\mathcal{P}}\psi) = 1$. There are two cases to consider. Either β is a successor ordinal or it is a limit ordinal. If it is a successor, then it follows that $\forall v \in \mathcal{V}^{<\alpha,\beta-1>} : v(\psi) = \frac{1}{2}$. By Lemma 1, $\mathcal{V}^{<\alpha,\beta>} \subseteq \mathcal{V}^{<\alpha,\beta-1>}$, so $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\psi) = \frac{1}{2}$. Therefore, $v^{<\alpha,\beta+1>}(O^{\mathcal{P}}\psi) = 1$. If it is a limit ordinal, it follows that $\forall v \in \bigcap\{\mathcal{V}^{<\alpha,\eta>} : \eta < \beta\}$: $v(\psi) = \frac{1}{2}$. By Lemma 1, $\mathcal{V}^{<\alpha,\beta>} \subseteq \bigcap\{\mathcal{V}^{<\alpha,\eta>} : \eta < \beta\}$, so $\forall v \in \mathcal{V}^{<\alpha,\beta>} : v(\psi) = \frac{1}{2}$. Therefore, $v^{<\alpha,\beta+1>}(O^{\mathcal{P}}\psi) = 1$. (The case for 0 is similar.)

Now that we've established that for each α, the micro-construction starting from $v^{<\alpha,0>}$ is monotonic and thus reaches a fixed-point, $v^{<\alpha,\gamma>}$, we move on to item 3: if $v^{<\alpha,\gamma>}(\phi) = 1(0)$, then $v^{<\alpha+1,0>}(\phi) = 1(0)$.

- For atomic sentences and logical sentences we reason as before.
- Once more the only interesting cases are those involving $O^{\mathcal{H}}$ and $O^{\mathcal{P}}$. But note that $O^{\mathcal{P}}$-sentences do not change their value in the transition from $v^{<\alpha,\gamma>}$ to $v^{<\alpha+1,0>}$, so we only need to look at sentences of the form $O^{\mathcal{H}}\psi$ now.

  – Let's assume that $v^{<\alpha,\gamma>}(O^{\mathcal{H}}\psi) = 0$. Given that $O^{\mathcal{H}}$-sentences do not change their value within micro-constructions, $v^{<\alpha,0>}(O^{\mathcal{H}}\psi) = 0$. α is either a successor ordinal or a limit ordinal. If α is a successor, let $v^{<\alpha-1,\delta>}$ be the fixed-point of the micro-construction starting from $v^{<\alpha-1,0>}$. Then $\forall v \in \mathcal{V}^{<\alpha-1,\delta>} : v(\psi) = i$ for $i \in \{1, 0, \frac{1}{2}\}$. By Lemma 1, $\mathcal{V}^{<\alpha,\gamma>} \subseteq \mathcal{V}^{<\alpha-1,\delta>}$, so $\forall v \in \mathcal{V}^{<\alpha,\gamma>} : v(\psi) = i$ for $i \in \{1, 0, \frac{1}{2}\}$. Hence, $v^{<\alpha+1,0>}(O^{\mathcal{H}}\psi) = 0$. (If α is a limit ordinal the reasoning is similar.)

  – Let's assume that $v^{<\alpha,\gamma>}(O^{\mathcal{H}}\psi) = 1$. We need to prove that $v^{<\alpha+1,0>}(O^{\mathcal{H}}\psi) = 1$. Here the decreasingness of the set of $\mathcal{V}$s is not sufficient to show this. What we need is to guarantee that the behaviour of ψ remains constant as we move from one stage to the next. Now, either α is a successor ordinal or it is a limit ordinal. In both cases, $\psi \in \mathcal{L}^{Tr}$, so ψ doesn't contain any occurrences of $O^{\mathcal{H}}$ or of $O^{\mathcal{P}}$. Hence, the behaviour of ψ remains the same throughout the construction, and this means that $v^{<\alpha+1,0>}(O^{\mathcal{H}}\psi) = 1$.

□

References

1. Bacon, A. 2002. Can The Classical Logician Avoid The Revenge Paradoxes? *Philosophical Review* 124:299–352.
2. Beall, J.C. 2006. True, false and paranormal. *Analysis* 66:2–14.
3. Beall, J.C., textitSpandrels of Truth. Oxford: Oxford University Press.

4. Billon, A. 2019. Paradoxical hypodoxes. *Synthese* 196:5205–5229.
5. Castaldo, L. 2021. Fixed-point models for paradoxical predicates. *The Australasian Journal of Logic* 18: 688-723.
6. Chihara, C. 1979. The Semantic Paradoxes: A Diagnostic Investigation. *Philosophical Review* 88: 590-618.
7. Cook. R. T. 2011. The No-No Paradox is a Paradox. *Australasian Journal of Philosophy* 88: 467-482.
8. Cook, R. T. 2022. Outline of an Intensional Theory of Truth, *Notre Dame Journal of Formal Logic* 63: 81-108.
9. Cook, R. T. and N. Tourville. 2016. Embracing the technicalities: Expressive completeness and revenge, *The Review of Symbolic Logic* 9: 325-358.
10. Cook, R. T. and N. Tourville. 2020. Embracing intensionality: Paradoxicality and semi-truth operators in fixed point models, *Logic Journal of the IGPL* 28: 747–770.
11. Eldridge-Smith, P. 2007. Paradoxes and Hypodoxes of Time Travel. In *Art and Time*, eds. Jan Lloyd Jones, Paul Campbell & Peter Wylie, 172–189. Australian Scholarly Publishing.
12. Field, H. 2008. *Saving Truth from Paradox*. New York: Oxford University Press.
13. Field, H. 2020. Properties, propositions and conditionals, *Australasian Philosophical Review* 4: 112–146.
14. Fitting, M. 1986. Notes on the mathematical aspects of Kripke's theory of truth, *Notre Dame Journal of Formal Logic* 27: 75-88.
15. Gallovich, C. 2022. Semantic supervenience in fixed-point models, unpublished manuscript.
16. Gupta, A. 1982. Truth and Paradox, *Journal of Philosophical Logic* 11: 1-60.
17. A. Gupta and N. Belnap. 1993. *The Revision Theory of Truth*. Cambridge, Mass: MIT Press.
18. Herzberger, H. 1982. Notes on Naive Semantics, *Journal of Philosophical Logic* 11: 243-268.
19. Kripke, S. 1975. Outline of a theory of truth, *Journal of Philosophy* 72: 690-716.
20. Murzi, J. and L. Rossi. 2020. Generalized Revenge, *Australasian Journal of Philosophy* 98: 153-177.
21. Picollo, L. 2020. Alethic Reference, *Journal of Philosophical Logic* 49: 417-438.
22. Ripley, D. 2012. Conservatively Extending Classical Logic with Transparent Truth, *Review of Sym- bolic Logic* 5:354—378.
23. Rosenblatt, L. 2020. Maximal Non-trivial Sets of Instances of Your Least Favorite Logical Principle, *The Journal of Philosophy* 111:30–54.
24. Rosenblatt, L. 2021. Classical Recapture and Maximality, *Philosophical Studies* 178:1951–1970, 2021.
25. Rosenblatt, L. 2022. Should the Non-Classical Logician be Embarrassed?, *Philosophy and Phenomenological Research* 104: 388–407.
26. Rosenblatt, L. 2023. Paradoxicality Without Paradox, *Erkenntnis* 88:1347–1366.
27. Rosenblatt, L. and C. Gallovich. 2022. Paradoxicality in Kripke's theory of truth, *Synthese* 200.
28. Rosenblatt, L. and D. Szmuc. 2014. On Pathological Truths, *The Review of Symbolic Logic* 7: 601-617.
29. Sorensen, R. 2003. *Vagueness and Contradiction*. New York: Oxford University Press.
30. Visser, A. 1984. Four valued semantics and the liar, *Journal of Philosophical Logic* 13: 181-212.
31. Yablo, S. 2003. New grounds for naive truth, in *Liars and Heaps*, ed. J.C. Beall. Oxford University Press.