How to Conquer the Liar
an informal exposition

Boris Čulina
Department of Mathematics
University of Applied Sciences Velika Gorica
Zagrebačka cesta 5, 10410 Velika Gorica, Croatia
e-mail: boris.culina@vvg.hr

Abstract. This article informally presents a solution to the paradoxes of truth and shows how the solution solves classical paradoxes (such as the original Liar) as well as paradoxes that were invented as counter-arguments for various proposed solutions to the paradoxes of truth ("revenges of the Liar"). Also, an erroneous critique of Kripke-Feferman axiomatic theory of truth, which is present in contemporary literature, is pointed out.

keywords: paradoxes of truth, the truth predicate, the logical concept of truth, revenge of the Liar, the strong Kleene three-valued semantics, the largest intrinsic fixed point, Kripke-Feferman theory of truth

The concept of truth has various aspects and is a frequent subject of philosophical discussions. Philosophical theories usually consider the concept of truth from a wide perspective. They are concerned with questions such as – Is there any connection between the truth and the world? And, if there is – What is the nature of the connection? Contrary to these theories, the analysis of the paradoxes of truth is of a logical nature. It deals with the internal semantic structure of language, the mutual semantic connection of sentences, above all the connection of sentences that speak about the truth of other sentences and sentences whose truth they speak about. Truth paradoxes show that there is a problem in our basic understanding of the language and they are a test for any proposed solution. Thereby, it is important to make a distinction between the normative and analytical aspect of the solution. The former tries to ensure that paradoxes will not emerge. The latter attempts to explain why paradoxes arise and to construct a solution based on that explanation. Of course, the practical aspect of the solution is also important. It tries to ensure a good framework for logical foundations of knowledge, for related problems in Artificial Intelligence and for the analysis of the natural language.

In the twentieth century, two solutions stood out, Tarski's [Tar33] and Kripke's [Kri75] solution. Tarski's analysis emphasized the T-scheme as the basic intuitive principle for the concept of truth, but it also showed its inconsistency with the classical logic. Tarski's solution is to preserve the classical logic and to restrict the scheme: we can talk about the truth of sentences of a language only inside another essentially richer metalanguage. This solution is in harmony with the idea of reflexivity of thinking and it has become very fertile for mathematics and science in general. But it has normative nature – truth paradoxes are avoided in a way that in such frame we cannot even express paradoxical sentences. It is also too restrictive because, for the same reason we cannot express a situation in which there is a circular reference of some sentences to other sentences, no matter how common and harmless such a situation may be. Kripke showed that there is no natural restriction to the T-scheme and we have to accept it. But then we must also accept the riskiness of sentences – the possibility that under some circumstances a sentence does not have the classical truth value but it is undetermined. This leads to languages with three-valued semantics. Kripke did not give any definite model, but he gave a theoretical framework for investigations of various models – each fixed point in each three-valued semantics can be a model for the logical concept of truth. However, as with Tarski, the proposed solutions are normative – we can express the paradoxical sentences, but we escape a contradiction by declaring them undetermined.

Kripke took some steps in the direction of finding an analytical solution. He preferred the strong Kleene three-valued semantics for which he wrote it was "appropriate" but did not explain why it was appropriate. One reason for such a choice is probably that Kripke finds paradoxical sentences meaningful. This eliminates the weak Kleene three valued semantics which corresponds to the idea that paradoxical sentences are meaningless, and thus indeterminate. Another reason could be that the strong Kleene three valued semantics has the so-called investigative interpretation. According to this interpretation, this semantics corresponds to the classical determination of truth, whereby all sentences that do not have an already determined value are temporarily considered indeterminate. When we determine the truth value of these sentences, then we can also determine the truth value of the sentences that are composed of them. Kripke supplemented this investigative interpretation with an intuition about learning the concept of truth. That intuition deals with how we can teach someone who is a competent user of an initial language (without the truth predicate symbol "T") to use sentences that contain the truth predicate symbol "T". That person knows which sentences of the initial language are true and which are not. We give her the rule to assign the T attribute to the former and deny that attribute to the latter. In that way, some new sentences that contain the predicate of truth, and which were indeterminate until then, become determinate. So the person gets a new set of true and false sentences with which he continues the procedure. This intuition leads directly to the smallest fixed point of strong Kleene semantics as an analytically acceptable model for the logical concept of truth.

In [Čul01], a different analytical solution to the problem of the paradoxes of truth is given. In the labyrinth of literature on the paradoxes of truth [BGR20], this solution is positioned as the classical semantic closure of the largest intrinsic fixed point of the strong Kleene three valued semantics whith a specific interpretation. I will briefly and informally sketch this solution here and then show how it solves classical paradoxes (such as the original Liar) as well as paradoxes that have been invented as counter-arguments for other solutions to the problem of the paradox of truth ("revenges of the Liar"). All these informal considerations can be formalized by the means developed in the cited article.

1 An analysis of the paradoxes of truth

Besides a syntactic structure and an internal semantic structure, a language has an external semantic structure too, a connection between language forms and the subject of the language. The connection is based on certain external assumptions on the language use, one of which is that every atomic sentence is either true or false. These assumptions have grown from everyday use of language where we are accustomed to their fulfilment, but there are situations when they are not fulfilled. The Liar paradox and other paradoxes of truth are witnesses of such situations. The paradoxes are the results of a tension between implicitly accepted assumptions on the language and their non-fulfilment. Let's consider the sentence L (the Liar):

L: $\overline{L}$ is a false sentence. (or "This sentence is false.")

Using the usual understanding of the language, to investigate truth of L we must investigate what it says. But it says precisely about its own truth, and in a contradictory way. If we asume it is true, then it is true what it says – that it is false. But if we assume it is false, then it is false what it says, that it is false, so it is true. Therefore, it is a self-contradictory sentence. What is disturbing is the paradoxical situation that we cannot determine its truth value. The same paradoxicality, but without contradiction, emerges in the investigation of the following sentence I (the Truth-teller):

I: $\overline{I}$ is a true sentence. (or "This sentence is true.")

Contrary to the Liar to which we can't associate any truth value, to this sentence we can associate the truth as well as the falsehood with equal mistrust. There are no additional specifications which would make a choice between the two possibilities, not because we haven't enough knowledge, but principally. Therefore, we can't associate a truth value to these sentence, either.

I will begin the analysis with a basic observation that the previous sentences are meaningful, because we understand well what they say, even more, we used that in the unsuccessful determination of their truth values. However, they witness the failure of the classical procedure for the truth determination in some "extreme" situations. Paradoxicality emerges from a confrontation of the implicit assumption of the success of the procedure and the discovery of the failure. According to the classical procedure, the examination of the truth of a sentence is reduced to the examination of the truth of the sentences from which it is constructed. Thus, for example, the examination of the truth of a sentence of the form φ *or* ψ is reduced to the examination of the truth of the sentences φ and ψ. The reduction is performed according to the truth conditions for the logical connective *or*: φ *or* ψ is true when at least one of the sentences φ and ψ is true, and false when both φ and ψ are false sentences. Likewise, a sentence of the form $\forall x P(x)$ is true when the sentence $P(x)$ is true for every valuation of the variable x, and it is false when $P(x)$ is false for at least one valuation of the variable x. Thus, the examination of the truth of a sentence comes down to the examination of the truth of the sentences from which it is constructed, (if these sentences contain free variables, then we must look at all valuations of these variables). Examining the truth of these sentences is in the same way reduced to examining the truth of the sentences from which they are constructed, etc. All these chains of problem reduction end in atomic sentences. In common situations a language doesn't talk about the truth values of its own sentences, so the truth values of its atomic sentences don't depend on the truth values of some other sentences. To investigate their truth values we must investigate external reality they are talking about. The assumption of the classical language is that every atomic sentence has a definite truth value, so the procedure of determination of the truth value of every sentence also gives a definite truth value, True or False. Formally, it is secured by the recursion principle which says that there is a unique function from sentences to truth values, which obeys the classical truth conditions and its values on atomic sentences are identical to externally given truth values. But the classical situation can be, and is, disrupted when atomic sentences use the truth predicate symbol to speak of the truth of other sentences of the language. These are sentences of the form $\mathrm{T}(\overline{\varphi})$, where "T" is the truth predicate symbol, and $\overline{\varphi}$ is the name of a sentence φ of the language. The truth predicate symbol is part of the logical vocabulary of the language, like connectives, quantifiers and the predicate symbol of equality. The only difference is in universality. Only a language that has its own sentences in the domain of its interpretation (possibly through coding) can have a symbol of its own truth predicate. As we set, for example, the internal language conditions on the truth of the sentence φ *or* ψ in relation to the truth of the sentences φ and ψ, so we also set the internal truth condition on the sentence $\mathrm{T}(\ulcorner\varphi\urcorner)$. According to the meaning of the truth predicate symbol, this sentence is true when φ is true and false when φ is false. Now the reduction chains of examining the truth of a sentence do not necessarily end in atomic sentences. If any of the atomic sentences in the chain is of the form $\mathrm{T}(\ulcorner\varphi\urcorner)$ then the chain, under the conditions of truth for the truth predicate, is extended to the examination of the truth of the sentence φ. So it can happen that some of the chains of reduction never stop. This does not necessarily mean the failure to determine the truth of the initial sentence, as happened with the Liar, where we cannot make an evaluation along the corresponding chain that satisfies the truth conditions, or with the Truth-teller, where more evaluations meet the truth conditions. Take for example the following sentence:

$$\mathit{Log}:\ \mathrm{T}(\overline{\mathit{Log}})\ \mathit{or}\ \mathrm{T}(\overline{\mathit{not}\,\mathit{Log}})\ \text{(or "This sentence is true or false")}$$

If *Log* were false then, by the truth conditions, $\mathrm{T}(\overline{\mathit{not}\,\mathit{Log}})$ would be false, *notLog* would be false too, and finally *Log* would be true. Therefore, such truth valuation is impossible. But if we assume that *Log* is true, the truth conditions generate a unique consistent valuation. Therefore, the truth determination procedure gives the unique answer – that *Log* is true. Although circularity is present, the classical procedure of determining truth has been successfully completed – it has provided a unique solution.

2 Proposed solution

The paradoxes of truth stem precisely from the fact that the classical procedure of determining truth value does not always have to give a classically assumed (and expected) answer. Such an assumption is an unjustified generalization from common situations to all situations. The awareness of that transforms paradoxes of truth to normal situations inherent to the classical procedure. It also follows from this diagnosis how we can intervene in the semantic structure of the language with its own truth predicate by incorporating the failures of the classical procedure into it. We can preserve the classical procedure of the truth determination and consequently the internal semantic structure of language, but we must reject universality of the assumption of its success. The rejection doesn't change the meaning of the classical conditions on the truth function, because they are stated in a way independent of the assumption that the function is everywhere defined. Their functioning in the new situation is illustrated in the following sentence:

$L \ or \ 0 = 0$

On the classical condition for the connective *or*, this sentence is true precisely when at least one of the basic sentences is true. Because $\overline{0 = 0}$ is true consequently the total sentence is true, regardless of the fact that L hasn't the truth value. Equally, if we apply the truth value condition on the connective *and* to the sentence

$L \ and \ 0 = 0$

the truth value will not be determined. Namely, for the sentence to be true both basic sentences must be true, and it is not fulfilled. For it to be false at least one basic sentence must be false and this also is not fulfilled. So, non-existence of the truth value for L leads to non-existence of the truth value for the whole sentence.

The classical truth value conditions specify the truth value of a compound sentence in terms of truth values of its direct components regardless whether they have truth values or not. The lack of some truth value may lead, but does not have to, to the lack of the truth value of the compound sentence. It is completely determined by the classical meaning of the construction of a sentence and by the basic assumption that all sentences are considered meaningful regardless of the truth value. Therefore, some sentences, although meaningful, valued by the classical conditions have not the classical truth value, because the conditions do not give them a unique truth value. This leads to the *partial two-valued semantics of the language*. Where the procedure gives a unique truth value, truth or falsehood, we accept it, where it fails because it does not give any truth value or permits both values, the sentence remains without the truth value. This kind of semantics can be described as the *three-valued semantics of the language* – simply the failure of the procedure will be declared as a third value (*Undetermined*). It has not any additional philosophical charge. It is only a convenient technical tool for the description. Thus, the classical procedure of determining truth unambiguously generates the three-valued semantics of language: when it is successful, it gives a classical value, and when it is unsuccessful, it gives the value *Undetermined*, as a sign of its failure. Analysing the propagation of failure through the structure of sentences, we can easily see that this three-valued semantics of the language is precisely the strong Kleene three-valued semantics. It is not the result of an investigative procedure, as with Kripke, but the result of the classical truth determination procedure accompanied by the propagation of its own failure. An analysis of the circularities or some other forms of infinite reduction in the determination of the classical truth value gives the criterion of when the classical procedure succeeds and when it fails, when the sentences will have the classical truth value and when they will not. It turns out [Čul01] that the truth values of sentences thus obtained give exactly the largest intrinsic fixed point of the strong Kleene three-valued semantics. While the Kripke procedure is a bottom-up procedure and leads to the smallest fixed point, this procedure is a top-down procedure and leads to the largest intrinsic fixed point. In that way, the argumentation is given for that choice among all fixed points of all monotone three valued semantics for the model of the logical concept of truth. This semantics is completely determined by the classical semantics of the language and its failures in determining the truth values of sentences.

So, for now we have two semantics of the language. We have *classical* or *naive semantics* in which paradoxes occur because it assumes that each sentence is true or false, i.e. it assumes that the process of determining truth always gives an unambiguous answer. And we have its repair to the partial two-valued semantics of the language, i.e. to the three-valued semantics of the language, which accepts the possibility of failure of the classical procedure of determining truth. I will call this semantics the *primary semantics* of the language. However, to remain on the three-valued semantics would mean that the logic would not be classical, the one we are accustomed to. Concerning the truth predicate itself, it would imply the preservation of its classical logical sense in the two-valued part of the language extended by the "silence" in the part where the classical procedure fails. Although in a meta-description, $\mathrm{T}(\bar{\varphi})$ has the same truth value (in the three-valued semantic frame) as φ, that semantics is no longer the initial classical semantics (although it extends it) nor it can be expressed in the language itself – the language is silent about the third value, or better said, the third value is the reflection in the metalanguage of the silence in the language. So the expressive power of the language is weak. For example, the Liar is undetermined. Although we have easily said it in metalanguage we cannot express in the language itself, because, as it has already been said (in metalanguage), the Liar is undetermined. Not only that the "zone of silence" is unsatisfactory because of the previously stated reasons (it leads to the three-valued logic, it loses the primary sense of the truth predicate and it weakens the expressive power of the language), but it can be overcome by a natural *additional valuation* of the sentences which emerges from recognising the failure of the classical procedure. This point will be illustrated on the example of the Liar. On the intuitive level of thinking, by recognising the Liar is not true nor false we state that it is undetermined. So, it is not true what it claims – that it is false. Therefore, the Liar is false. But this does not lead to restoring of the contradiction because a *semantic shift* has happened from the primary partial two-valued semantics (or three-valued semantics) toward its two-valued description, which merely extends it in the part where it is not determined. Namely, the Liar talks of its own truth in the frame of the primary semantics, while the last valuation is in the frame of another semantics, which I will term the *final semantics* of the language. The falsehood of the Liar in the final semantics doesn't mean that it is true what it says (that it is false) because the semantic frame is not the same. It means that it is false (in the final semantics) what the Liar talks of its own primary semantics (that it is false in the primary semantics), because it is undetermined in the primary semantics. So, not only have we gained a contradiction in the naive semantics, i.e. the third value in the primary semantic, but we also have gained additional information about the Liar.

It is easy to legalize this intuition. Sentences of the language will always have the same meaning, but the language will have two valuation schemes – primary and final truth valuation. In both semantics the meaning of the truth predicate is the same: $\mathrm{T}(\bar{\varphi})$ means that φ is true in the primary semantics. But the valuation of the truth of that atomic sentence $\mathrm{T}(\bar{\varphi})$ is different. While in the primary semantics the truth conditions for $\mathrm{T}(\bar{\varphi})$ are classical (the truth of $\mathrm{T}(\bar{\varphi})$ means the truth of φ, the falsehood of $\mathrm{T}(\bar{\varphi})$ means the falsehood of φ) (and consequently $\mathrm{T}(\bar{\varphi})$ is undetermined just when φ is undetermined), in the final semantics it is not so. In it the truth of $\mathrm{T}(\bar{\varphi})$ means that φ is true in the primary semantics, and falsehood of $\mathrm{T}(\bar{\varphi})$ means that φ is not true in the primary semantics. It does not mean that it is false in the primary semantics, but that it is false or undetermined. So, formally looking, in the final semantics $\mathrm{T}(\bar{\varphi})$ inherits truth from the primary semantics, while other values transform to falsehood. That is why we say that this semantics is the *classical semantic closure* of the primary semantics, or in full terminology, the classical semantic closure of the largest intrinsic fixed point of the strong Kleene three-valued semantics. Due to the monotonicity of the primary semantics this means that the final semantics supplements the primary semantics in the area of its silence. If a sentence in the primary semantics has a classical value (True or False), it will have that value in the final semantics as well. If a sentence is indeterminate in the primary semantics (a paradoxical sentence) then it will have a classical truth value in the final semantics that just carries information about its indeterminacy in the primary semantics. Therefore, the final semantics is the classical two-valued semantics of the language that has for its subject precisely the primary semantics of the language which it extends in the part where it is silent using the informations about the silence.

We can see best that this is a right and a complete description of the valuation in the primary semantics by introducing predicates for other truth values in the primary valuation:

$\mathrm{F}(\bar{\varphi})$ (= φ is false in the primary semantics) $\leftrightarrow \mathrm{T}(\overline{\mathit{not}\,\varphi})$

$\mathrm{U}(\bar{\varphi})$ (= φ is undetermined in the primary semantics) $\leftrightarrow \mathit{not}\,\mathrm{T}(\bar{\varphi}) \wedge$

not $\mathrm{F}(\bar{\varphi})$

According to the truth value of the sentence φ in the primary semantics we determine which of the previous sentences are true and which are false. For example, if φ is false in the primary semantics then $\mathrm{F}(\bar{\varphi})$ is true while others ($\mathrm{T}(\bar{\varphi})$ and $\mathrm{U}(\bar{\varphi})$) are false.

Once the final two-valued valuations of atomic sentences are determined in this way, valuation of every sentence is determined by means of the classical conditions and the principle of recursion. This valuation not only preserves the primary logical meaning of the truth predicate (as the truth predicate of the primary semantics) but it also coincides with the primary valuation where it is determined. Namely, if $\mathrm{T}(\bar{\varphi})$ is true in the primary semantics then φ is true in the primary semantics, so $\mathrm{T}(\bar{\varphi})$ is true in the final semantics. If $\mathrm{T}(\bar{\varphi})$ is false in the primary semantics then φ is false in the primary semantics, so $\mathrm{T}(\bar{\varphi})$ is false in the final semantics. Since the truth conditions for compound sentences are the same in both semantics this coincidence spreads through all sentences which have determined value in the primary valuation. Therefore $\mathrm{T}(\bar{\varphi}) \to \varphi$ and $\mathrm{F}(\bar{\varphi}) \to \mathit{not}\, \varphi$ are true sentences in the final semantics.

The final semantics describes the truth predicate of the primary semantics, not its own truth predicate. Thus, according to Kripke, “The ghost of the Tarski hierarchy is still with us.”. I do not think so. The concept of truth in the final semantics is not a logical concept of truth. It is equal to the concept of truth in other sciences. Just as, for example, a language describes car engines, here the final semantics describes the truth predicate of the primary semantics. Of course, as in the languages of mechanical engineering, the question of the truth of sentences can be asked here, as well as the truth of sentences that speak of the truths of sentences that speak of the truth predicate of the primary semantics, etc. But this is a different type of problem than the problem of paradoxical sentences. In [Čul20] it is shown how this infinite regression naturally stops in the corresponding metalanguage. Metalanguage is used here as well to describe the primary and the final semantics of the language with its own truth predicate of the primary semantics.

3 Conquering the Liar

Having in mind this double semantics of the language (triple, if we also count the classical naive semantics), we can easily solve all truth paradoxes. On an intuitive level we have already done it for the Liar:

L: $\mathrm{F}(\overline{L})$ ("This sentence is false.")

The form of the solution is always the same. A paradox in the classical thinking means that the truth value of a sentence is undetermined in the primary semantics. But, then it becomes an information in the final semantics with which we can conclude the truth value of the sentence in the final semantics. To make it easier to track solutions to other paradoxes, I will sometimes distinguish by appropriate prefixes what the truth valuation is about: I will put prefix "p" for the primary semantics and prefix "f" for the final semantics. In that way we will distinguish for example "f-falsehood" and "p-falsehood".

The Strengthened Liar is the "revenge of the Liar" for solutions that seek a way out in truth value gaps, i.e. in the introduction of a third value – indeterminate:

SL: *not* $\mathrm{T}(\overline{SL})$ ("This sentence is not true.")

In the classical semantics it leads to a contradiction in the same way as the Liar, because there "not to be true" is the same as "to be false". The paradox is used as an argument against the third value in the following way (e.g. in [Bur79]). If we accept that The Strengthened Liar takes on the value *Undetermined*, it means that what it is saying is true – that it is not true (but indeterminate) – and so the contradiction is renewed. However, the last step is wrong because a semantic shift has occurred! The conclusion that The Strengthened Liar is undetermined is the conclusion in the final semantics. So when we say that what he says is true, this is the concept of truth of the final semantics, while the concept of truth he mentions is the concept of truth of the primary semantics. So the truth of the final semantics is that The Strengthened Liar is not true in the primary semantics. Or, using prefixes, we can also state this in the following way. Recognising a failure of the classical procedure, we continue to think in the final semantics and state that it is p-undetermined. So, it is not p-true. But, it claims just that, so it is f-true. Therefore, we conclude that the Strengthened Liar is p-undetermined and f-true. It is interesting that the whole argumentation can be done directly in the final semantics, not indirectly by stating the failure of the classical procedure. The argumentation is the following. If SL were f-false, then it would be f-false what it said – that it is not p-true. So, it would be p-true. But, it means (because the final semantics extends the primary one) that it would be f-true and it is a contradiction with the assumption. So, it is f-true. This statement does not lead to a contradiction but to an additional information. Namely, it follows that what it talks about is f-true – that it is not p-true. So, it is p-false or p-undetermined. If it were p-false it would be f-false too, and this is a contradiction. So, it is p-undetermined.

Note that, although the Liar and the Strengthened Liar are both p-undetermined, the latter is f-true while the former is f-false.

In [Bur79], Burge also introduces the following the revenge of the Liar for truth value gaps solutions:

BL: $\mathrm{F}(\overline{SL})$ *or* $\mathrm{U}(\overline{SL})$ ("This sentence is false or undetermined.")

When we consider it in the classical semantics, if it were true then it would be false or indeterminate, which is a contradiction. If it were false, then it would be true – again a contradiction. So, again we make a semantic shift and in the final semantics we conclude that it is indeterminate. This means that in the final semantics it is true. Or, if we express ourselves with prefixes, that sentence is p-indeterminate and f-true.

The semantic shift in argumentation is best seen in the following variant, the so-called Metaliar:

1. The sentence on line 1 is not true.

2. The sentence on line 1 is not true.

The sentence on line 1 is The Strengthened Liar so it is indeterminate. If we understand the second sentence as reflection on the first sentence, which we have determined to be indeterminate, then the second sentence is true. So it turns out that one and the same sentence is both indeterminate and true. In [Gai92], Gaifman uses this example to motivate the association of truth values not with sentences as sentence types but with sentences as sentence tokens. Thus, Gaifman solves the paradox by separating the same sentence type into two tokens of which the first is indeterminate and the second true. In my approach, it is precisely the separation of the primary and the final semantics of the same sentence. In the 1st line it gets the indeterminate value in the primary semantics, while in the 2nd, by reflection on the primary semantics, it gets the value *True* in the final semantics.

In [Sky84], Skyrms introduced the Intensional Liar, to point out the intensional character of the Liar. Namely, if in The Strengthened Liar

(1): (1) is not true.

we replace (1) with the standard name of the sentence denoted by that sign, we get the sentence

"(1) is not true" is not true.

While sentence (1) is indeterminate, this harmless substitution seems to have given us the sentence which is not indeterminate but true, for the sentence she speaks of is indeterminate, therefore it is not true. But here, too, there has been a semantic shift in truth evaluation that we can explain with prefixes:

" " (1) is not true" is not p-true" is f-true

4 Conquering the companions of the Liar

In the same way, paradoxes that have a different type of failure of the classical procedure, such as the Yablo paradox, are solved [Yab93]. Consider the following infinite set of sentences $(i), i \in N$:

(i) For all $k > i$ (k) is not true.

If the sentence (i) were true, then all the following sentences would not be true. But that would mean on the one hand that (i + 1) is not true, and on the other hand, since all the sentences after it are not true, that (i + 1) is true. So all the above sentences are not true. But if we look what they claim entails that they are all true. This contradiction in the classical semantics turns into a true claim of the final semantics that all these sentences are p-indeterminate. From what they say about their primary semantics, as with the the Strengthened Liar, it follows that they are all f-true.

Note that in Yablo's paradoxical sentences the chains of reduction in examining their truths are different than in the previously examined paradoxical sentences. There is no circularity in the chains, i.e. they are not composed of the same repetitive cycle.

That the solution of the problem of the paradox of truth presented here is not related to negation will be illustrated by the example of Curry's paradox [Cur42]:

C: $\mathrm{T}(\bar{C}) \to l$ ("If this sentence is true then l")

where l is any false statement. On the intuitive level, if C were false then its antecedent $\mathrm{T}(\bar{C})$ is true, and so is C itself, and it is a contradiction. If C was true then the whole conditional (C) and its antecedent $\mathrm{T}(\bar{C})$ would be true, and so the consequent l would be true, which is impossible with the choice of l as a false sentence. Therefore we conclude in the final semantics that C is p-undetermined, and so it is f-true (because the antecedent is f-false).

All the previous paradoxical sentences led to contradictions in the classical semantics. Thus, in the final semantics, we came to the conclusion that they are indeterminate in the primary semantics, from which we further determined their truth value in the final semantics. But we could also analyse them directly in the final semantics, as was done with the Strengthened Liar. There, the contradiction would turn into a positive classical two-valid argumentation by which we would determine its truth value in both the primary and the final semantics. However, the situation is different with paradoxes which do not lead to contradiction, which permit more valuations, like the Truthteller. Its analyses gives that it is p-undetermined. It implies that it is not p-true which means that (I : $\mathrm{T}(\bar{I})$) it is not I. So, I is f-false. However, although the conclusion is formulated in the final semantics, thinking alone cannot be formulated in it because it involves the analysis of the corresponding chains of reduction. Of course, if we enrich the language with the description of such chains and their truth valuations then it is possible to translate the whole intuitive argumentation.

In [Gup82], Gupta gave several arguments against Kripke's fixed points. The solution presented here is a semantic closure of the largest intrinsic fixed point of the strong Kleene three-valued semantics, so this critique also applies to it.

One of Gupta's criticisms, which is already present in the literature, is that not all classical laws of logic are valid in fixed points. Eg. for a language containing the Liar, the sentence $\forall x$ *not* ($\mathrm{T}(x)$ *and not* $\mathrm{T}(x)$) is undetermined in each fixed point of the strong Kleene three-valued semantics (if we choose the Liar for x, we get the indeterminate sentence). But since the analysis of paradoxes cannot avoid the presence of sentences that have no classical truth value, this naturally leads to a three-valued language for which we cannot expect the logical laws of a two-valued language to apply. The strong Kleene three-valued semantics is maximally adapted to two-valued logic: the logical truths of two-valued logic are always true in it when they are determinate. However, the transition to final semantics definitely solves this problem because that semantics is two-valued.

A somewhat more inconvenient situation is that $\forall x$ *not* ($\mathrm{T}(x)$ *and not* $\mathrm{T}(x)$), like other logical laws, is not true in the smallest fixed point even when there is no the Liar or the Truth-teller, i.e. when there are no paradoxical sentences or sentences that take on different values at various fixed points. Namely, then the stated statement is not true for its own sake – in order to determine its truth, the truth of all sentences, including itself, must be examined. In this way it can be seen that it is an ungrounded sentence, i.e. indeterminate in the smallest fixed point. But in the greatest intrinsic fixed point, it is true. We can easily check this by trying to give it a classic valuation. Namely, in order to examine its truth, we must examine whether *not* ($\mathrm{T}(x)$ *and not* $\mathrm{T}(x)$) is valid for each sentence x. Since we assume that language has no paradoxical sentences, it is only necessary to examine whether this is true of the sentence itself. If this condition were false to it, then it would be false. But the negation of this condition would be true which entails that it is true, and so we get a contradiction. Thus it must be true (and this value does not lead to contradiction). It means that this Gupta's critique turns into an argument for the greatest intrinsic fixed point.

The second type of critique seeks to show that some quite intuitive considerations about the concept of truth are inconsistent with the fixed point model. Gupta constructed the following example in [Gup82] (Gupta's paradox). Let us have the following statements of persons A i B:

A says:

(a1) Two plus two is three. (false)

(a2) Snow is always black. (false)

(a3) Everything B says is true. (   )

(a4) Ten is a prime number. (false)

(a5) Something B says is not true. (   )

B says:

(b1) One plus one is two. (true)

(b2) My name is *B*. (true)

(b3) Snow is sometimes white. (true)

(b4) At most one thing *A* says is true. ( )

Sentences (a1), (a2), (a4), (b1), (b2) and (b3) are determinate in each fixed point. However, (a3) and (a5) " wait " (b4), and (b4) "waits" them and so those sentences remain indeterminate in the smallest fixed point. But on an intuitive level, it is quite easy for them to determine the classical truth value. Since (a3) and (a5) are contradictory, and all other statements of *A* are false, (b4) is true. But this means that (a3) is true and (a5) is false. However, this intuition coincides with the truth valuation in the greatest intrinsic fixed point. Thus this Gupta's critique also turns into an argument for the greatest intrinsic fixed point. In order to find an intuitive counterexample for it as well, Gupta replaces (a3) and (a5) with the following statements:

(a3*) (a3*) is true. ( )

(a5*) "(a3*) is not true" is true. ( )

Now at the largest fixed point, (a3 *) and (a5 *), and thus (b4), are indeterminate. Gupta considers that on an intuitive level (b4) is true, because at most one of (a3 *) and (a5 *) is true. But in the intuitive argument there is a semantic shift from the primary semantics to the final, because (b4) is a true statement in the final semantics.

5 An erroneous critique of Kripke-Feferman theory

The initial motivation for this article was to read two contemporary respectable books on formal theories, in the sections in which they talk about the Kripke-Feferman axiomatic theory of truth (KF). The models of this theory are the classical semantic closures of fixed points of strong Kleene semantics, and so the final semantics described in this paper, too.

In [Hor11] on page 127 is the following text:

> So far, it seems that KF is an attractive theory of truth. However, we now turn to properties of KF that disqualify it from ever becoming our favorite theory of truth.
>
> Corollary 70: $KF \vdash L \wedge \neg T(L)$, where L is the [strengthened] liar sentence.
>
> ...
>
> In other words, KF proves sentences that by its own lights are untrue. This does not look good. To prove sentences that by one's own lights are untrue seems a sure mark of philosophical unsoundness: It seems that KF falls prey to the strengthened liar problem.

In [BGR18] on page 76 is the following text:

> But on the properties of truth itself, KF also has some features some have found undesirable. One example (discussed at length in Horsten 2011) is that KF $\vdash \lambda \wedge \neg\, T\lambda$. Unlike FS, KF gives us a verdict on Liars. But it seems to then deny its own accuracy, as it first proves λ, and then denies its truth. This makes the truth predicate of KF awkward in some important ways.

Both quoted texts repeat KF's critique dating back to Reinhardt [Rei86], that axiomatic KF theory is not an acceptable theory of truth, which means that its models, classical semantic closures of fixed points, are not acceptable solutions to the concept of truth. The reason is that the theory proves both the Liar and that the Liar is not true. The error in this reasoning stems from the indistinguishability of the primary (fixed point) and the final (classical semantic closure of the fixed point) semantics. We have already seen that the Liar is true in the final semantics, so it is not awkward that KF $\vdash L$). It means that in the final semantics it is true that the Liar is not true in the primary semantics,so it is not awkward that KF $\vdash \neg T(\bar{L})$. These claims (in fact one claim) are not contradictory, because different concepts of truth are involved.

References

[BGR18] Jc Beall, Michael Glanzberg, and David Ripley. *Formal Theories of Truth*. Oxford: Oxford University Press, 2018.

[BGR20] J. Beall, M. Glanzberg, and D. Ripley. Liar Paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2020.

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

[Čul01] Boris Čulina. The concept of truth. *Synthese*, 126:339–360, 2001.

[Čul20] Boris Čulina. The synthetic concept of truth. unpublished, 2020.

[Cur42] Haskell B. Curry. The inconsistency of certain formal logics. *Journal of Symbolic Logic*, 7:115—-117, 1942.

[Gai92] Haim Gaifman. Pointers to truth. *Journal of Philosophy*, 89, 1992.

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

[Hor11] Leon Horsten. *The Tarskian Turn. Deflationism and Axiomatic Truth*. MIT Press, 2011.

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

[Rei86] William N. Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15(2):219–251, 1986.

[Sky84] Brian Skyrms. Intensional aspects of semantical self-reference. In Robert L. Martin, editor, *Recent Essays on Truth and the Liar Paradox*. Oxford University Press, 1984.

[Tar33] Alfred Tarski. Pojęcie prawdy w językach nauk dedukcyjnych. *Towarzystwo Naukowe Warszawskie*, 1933. njemački prijevod, “Der Wahrheitsbegriff in den formalisierten Sprachen”, Studia philosophica 1, 1935, 261–405.

[Yab93] Stephen Yablo. Paradox without self-reference. *Analysis*, 53:251–252, 1993.