DEFINABILITY AND UNDEFINABILITY OF TRUTH[1]

review paper

Tomislav KARAČIĆ

Master of Logic Programme
Institute for Logic, Language and Computation
University of Amsterdam
e-mail: tomislavkaracic.sc@gmail.com

Abstract

This thesis will focus on Tarski's work, so the Section 2 starts with him, giving an overview of the elements of his theory of truth, leading to a presentation of his theorem in 2.3. In the rest of the paper, proposals for solving the problem of internalization of the concept of truth and the paradoxes arising from this problem with some of the proposed solution are considered. Section 3. will quickly introduce the Liar paradox and once more explicitly state Tarski's solution, followed by the so called "revenge of the liar"; a liar-type sentence which cannot be avoided even by using Tarski's solution to the original liar. After that, Section 4. will introduce Kripke's Theory of Truth and his take on truth and solving the Liar paradox using paracomplete logic, while Section 5. will examine some further attempts in answering the aforementioned problems in the context of paraconsistent logic.

Key words

truth, Tarski's theorem, semantical paradoxes, liar, formal semantics

[1] Ovaj rad pisan je kao završni rad na Odsjeku za filozofiju Hrvatskih studija u Zagrebu za potrebe ostvarivanja prvostupništva na studiju filozofije.

1. Tarski's Theory of Truth

1.1. Importance of Tarski's Work for Semantics

During the early 20th century, among the philosophers who were driven in their work by the virtues promoted by the empiricist project, it was not uncommon to approach the concept of truth with a high dose of scepticism [13, p. 23]. The champions of such a way of doing philosophy were without doubt the positivists of the Vienna Circle. Keeping in mind their reluctance to accept or deal with metaphysical notions which could not be grounded in hard science, it is not surprising that a concept which has a long history of dubiousness, in the form of logical antinomies as well as well as supposing a link between language and the world around us, is regarded with suspicion. This suspicion was obvious towards any semantic notion, as well as semantics itself, and the reason is clear from this quote from Carnap:

> While many philosophers today urge the construction of a system of semantics, others, especially, among my fellow empiricists, are rather sceptical. They seem to think that pragmatics - as a theory of the use of language -is unobjectionable, along with syntax a purely formal analysis; but semantics arouses suspicions. They are afraid that a discussions of propositions -as distinguished from sentences expressing them- and of truth- as distinguished from confirmation by observations - will open the back door to speculative metaphysics, which was put out at the front door. [4]

In this respect, the project of the Vienna Circle was closely tied with that of the Lvov-Warsaw school, which, too, wanted to develop scientific philosophy based on mathematical logic as a paradigm of rational thought. Tarski, like Carnap, wrote about scepticism towards semantics mentioning antinomies as one of the reasons for such caution:

> Concepts from the domain of semantics have traditionally played a prominent part in the discussions of philosophers,

> logicians and philologists. Nevertheless they have long been regarded with a certain scepticism. From the historical point of view this scepti- cism was well founded; for, although the content of the semantical concepts, as they occur in colloquial language, is clear enough, yet all attempts to characterize this content more precisely have failed, and various discussions in which these concepts appeared and which were based on quite plausible and seemingly evident premises, had often led to paradoxes and antinomies. [23]

Tarski's work, as well as Carnap's and Gödel's for example, lead to the dissolution of this scepticism towards semantics and thanks to them, the importance of semantics in contemporary logic is unquestionable [25, p. 4]. When it comes to the notion of truth, Tarski's project was driven by an idea of demonstrating that the concept of truth could be defined in terms of other notations whose scientific examination was possible [10, p. 38] [6, p.535]. Also, this kind of general scepticism towards the concept of truth is a source of the problem of priority with regards to proving the Undefinability of truth theorem, which is covered here in 2.3.1.

It is important to note that Tarski did not define the notion of truth as commonly understood, or in other words he did not define truth in everyday sense of the word [22, p. 153]. Rather, he defined a notion of truth for a large class of formalized languages L, were truth is defined in L and whose definition is applicable to the sentences of L. Although Tarski writes about truth in formal systems and explicitly states at the beginning of his seminal 1933. paper that he wishes to give a materially adequate and formally correct definition of a term 'true sentence' [22, p. 187], his goal in constructing a theory of truth is indeed philosophically deeper in a sense that the result he wishes to obtain is constructing a materially accurate and formally con- sistent definition of the classical notion of truth (or being true) where 'true' means 'corresponding with reality' [12, p. 285]. Nevertheless, Tarski early on expresses great doubt about the possibility of defining truth for colloquial languages and states that he shall focus on formalized ones instead. More- over, he argues that, since colloquial languages are semantically closed, it is impossible to create a satisfactory definition of 'true' [21, p. 537]. This is due to the

universality of those kind of languages [7, p. 36, 58], which means that they contain all of the expressions we expect them to contain, the names of these expressions and semantic terms like 'true' that refer to the sentences of these languages.

For that reason definitions are needed of the terms "formal language", "formal correctness" and "material adequacy", so that Tarski's position can be properly understood.

A *formal language* is the one in which:

1. a description is given in structural terms of all the signs with which the expressions of a language are formed,
2. among all of the possible expressions, sentences are distinguished only by means of structural properties,
3. a structural description is given for sentences called axioms,
4. by rules of inference, structural operations are embodied which permit the transformation of one sentence into another.

Object language is a language L under discussion, while the metalanguage is a language M in which definitions of object language L are given. A met- alanguage M contains a copy of object language L, and M can talk about sentences and syntax of L. Furthermore, M contains a predicate True, which we read as 'is true in L'.

Formal correctness is a condition that a definition of the predicate True should be in a form For all x, True(x) if and only if ψ(x), where the predicate True never occurs in ψ.

Material adequacy, as a condition for a satisfactory definition of 'true', means that the sentences that satisfy ψ should be exactly those which we would regard as being true sentences of L, and that this fact should be provable from the axioms of the metalanguage. Material adequacy is what is also called *Convention T*.

1.2. Convention T

In the philosophical pursuit of the definition of truth, one can ask about a set of sentences if its members are true. Then,

one can try to identify a property that all true sentences have in common or a method which gives an answer for every sentence if it is true or false. But, these are not the questions we will deal with, since Tarski has a different project in mind. He is not trying to find a property which all true sentences have, but a condition which every theory of truth should fulfil to be an adequate theory [18, p. 82]. In other words, his goal is not directly to separate true sentences from false, but adequate theories from non-adequate theories.

Therefore, the purpose of the Convention T is to provide the condition upon which a sentence of L is true, or in other words the condition under which a sentence falls under the extension of 'true' and not to say if the sentence really is true. The extension of 'true' indicates the range of appli- cability of 'true' by naming the particular objects that it denotes, in this case, true sentences, while being true, on the other hand, is a property of sentences. So, the Convention T determines the criterion for sentences to be-come members of the extension of 'true'. That condition should also preserve our intuitions on truth. Hence, in cases where we known that a sentence is true, for example "2+3=5", any proposed condition must put "2+3=5" into the extension of 'true'. In cases such as the negation of "2+3=5", the condi- tions should put that sentence outside of the extension of 'true'. Moreover, non-sentences should also lie outside the extension of 'true'. The reason why Convention T merely provides a condition upon which a sentence of L is true is the existence of sentences for which we do not know if they fall under the extension of 'true'. For example, we do not know if Goldbach's Conjecture falls under 'true', but we are not required to know which sentences are true to have a definition of 'true'. Instead, we only need to explicitly say under which condition it would be true.

> **Convention T**. A proposed definition of truth for the language of arithmetic is adequate if it implies all sentences of the form:
> (T) $\lceil\psi\rceil$ is true if and only if ψ, when ψ is a sentence of the language of arithmetic and $\lceil\psi\rceil$ is its Gödel number; and it also implies $(\forall x)$(x is true $\rightarrow$ x is a sentence).

So, given any language L, Tarski argues that an adequate notion of truth for L would have to satisfy, for each sentence x, 'x' is true if and only if x [22, p. 187]. A paradigmatic example in English is the sentence 'Snow is white' which, by convention T, is true if and only if snow is white. The characteristic of correspondence, which is enveloped by such a notion of truth, finds its justification in, what Tarski would deem as an Aristotelian, definition of truth that establishes truth as correspondence formulated in Aristotel's dictum [12, p. 285]:

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

What Convention T states is that a predicate T is an adequate definition of truth if in the metalanguage all sentences which are obtained from 'T(x) if and only if p' by substituting for 'x' a structural-descriptive name of any sentence of L and for the 'p' the translation of this sentence into the meta- language. In other words, if we take the language of Gödel numbers as our object language and arithmatization as translation of that language to the metalanguage, then some predicate in the metalanguage T is adequate if and only if the following theorems can be proven in the that metalanguage:

a) All the theorems of the form T (x) → p, where x is a particular Gödel number and p is the translation of that Gödel number, and that
b) for all y, if T(y) can be proven in the metalanguage then y is a sentence of the object language.

With regards to defining truth, note the term *convention* in Convention T. It is important not to confuse Convention T with a definition of truth. What the Convention provides is a scheme which, when we substitute x with a name of a concrete expression and p with that expression, becomes a partial definition of truth. It is the conjunction of all of the partial definitions that is the general definition of truth. [10, p. 38]

1.2.1. Convention T and the Liar

Although the relation between the Liar paradox and Tarski's response to it will be thoroughly examined in the Section 3., a short introduction will be presented now, as it will shed light on the motivation for the establishment of the Convention T. Moreover, we will show in 2.3 how Tarski's theorem is proven by constructing a Liar-type sentence.

The Liar's Paradox is one of the oldest and most commonly known para- dox. It can be stated as:

1) This statement is not true.

The paradox arises since the sentence is true if it is not true and vice versa [8, p. 55]. Alfred Tarski has shown a way to solve this paradox, although in a specific situation, that is when a formal language does not contain its true predicate T [22, p. 262]. This is where we arrive at Tarski's Convention T.

So, since a central element of the Convention T is that the truth predicate T is not expressible in the object language L; instead, it is a predicate of the metalanguage ML which will take descriptions of the statements in the object language, it follows that (1) is not interpretable in the object language L because of the presence of the truth predicate T. In other words, it is not a well formed formula.

We can interpret it in ML as:

1)* ¬T ((1))

But again, this will not be a well formed formula since 1)* contains T and T can only be applied to sentences of L whose 1)* cannot be a member of because it contains T which is not expressible in L.
In this way, the paradoxes can be avoided with the use of metalanguages by showing that antinomies such as 1) are in fact not well formed formulas.

1.3. Tarski's Theorem and its Consequences

Tarski has shown the usefulness of metalanguages in dealing with semantic paradoxes, but the importance of the metalanguage goes beyond just dealing with the Liar. As Tarski demonstrates, in his theorem on the undefinability of truth predicate, without such a system of metalanguages, it is impossible to define truth for formal systems. Tarski's theorem shows that the resources of the metalanguage ML must go beyond the resources of the object language L, or in other words that it necessitates the use of a metalanguage ML that is expressively more powerful than the object language L for which it provides semantics [22, p. 254]. The proof is constructed by using the Diagonal lemma which states:

Diagonal Lemma. *Let Σ be a theory that represents every recursive function, and let A(x) be a formula in the language of Σ with just 'x' free. Then there is a sentence G such that Σ ⊢ tt ≡ A(⌈tt⌉), where ⌈tt⌉ is a Gödel number of a sentence G.*

G is a sentence that says of itself that it has whatever property that A(x) expresses [20, p. 200]. Through this lemma, the existence of self-referential sentences is mathematically provable.

Theorem 1. *Tarski's theorem*
(a) *In whatever way predicate T is defined in metalanguage ML, it will be possible to derive from it the negation of one of the sentences which were described in the condition for the convention T*;
(b) *assuming that the class of all provable sentences of the metatheory is consistent, it is impossible to construct an adequate definition of truth in the sense of convention T on the basis of the metalanguage ML.*

To prove the theorem, we will take T(z) to mean that the statement rep- resented by the Gödel number z is true, while we will use Q(y,y,z) to say that if we put 'y', the expression of the Gödel number y, into the statement repre- sented by the Gödel number y, then we will get the statement represented by the Gödel number z.

Using Gödel numbering, we can construct a sentence 2) (∃y)(¬T (z) ∧ Q(y, y, z)) which states that there exists some Gödel number y such that it is not the case that the statement which corresponds to the Gödel number z is true and that that same number is the arithmatization of the formula created by putting the expression of the Gödel number of 2) into that same formula.

If we assume that 2) is true, then the first part of the conjunction says that it is not the case that the proposition represented by z is true, while the second part says that the proposition represented by z is a result of putting the Gödel number of sentence 2) into the arithmetization of itself, which will result in the original formula. Therefore, it is not the case that T(z), while the translation of z in the metalanguage is true, which violates Convention T.

If we assume that 2) is false, we get a sentence 3) (∀z)(T (z)∨¬Q(y, y, z)). By instantiating z for the Gödel number of the sentence 2), Q(y,y,z) becomes true, so by disjunctive syllogism T(z) holds. Since z is the Gödel number of sentence 2) Convention T must hold, which contradicts our assumption that 2) is false.

Now, we will present a cleaner proof of Tarski's theorem [8, p. 63, 68].

> *Proof*. Let T^1 be a predicate which arithmetically codes the semantic notion of true.
>
> Let sb3 be a 3-placed function symbol, which is an arithmetical translation of syntactic substitution in a given formula, of a given expression for a particular variable.
>
> So, we can write sb(y, y, 13) to denote the Gödel number of a formula created by substitution in a formula with a Gödel number y, of the expression of the Gödel number y, for the variable with the Gödel number 13, 13 being the Gödel number of 'y', thus achieveing self-reference.
>
> Now, we can create a formula
>
> a) ¬Tsb(y, y, 13) with h being the Gödel number of a).

By substituting h for y we get
b) ¬Tsb(h, h, 13) with sb(h,h,13) being the Gödel number of b). Hence, b) states its own falsity and the Liar paradox arises.

□

The proof of Tarski's theorem shows that truth is undefinable in the object language and it also demonstrates that no consistent theory of truth for a language that one can formulate within that exact language implies the (T)-sentences. Moreover, such theories are inconsistent with the (T)- sentences.
Thus, to use 'true', we need to construct a hierarchy of languages, L_0, L_1, L_2,..., where each language L_{n+1} has a truth predicate T_{n+1} that can only be applied to the sentences of Lm, where m ≤ n holds. In this case, L_0 is what we call object language and L_{n+1} metalanguages.

However, there are several issues with such a hierarchy some of which will be covered in 3.1 and 4.1, while we will briefly mention one right away. The issue in question concerns the fact that by building higher and higher levels of this hierarchy of languages it is conceivable that at one point we will get to a language in which we formulate questions and state results of any kind. Intuitively, the best candidate for this language would be a natural language, for example Croatian [21, p. 537]. But, it is precisely this kind of language that Tarski deems inconsistent, so the problem is how to explain the link between truth as understood in formal languages and truth in everyday sense of the word. Thus, every time we go one step higher, the metalanguage we find ourselves in will never have a feature of universality. Instead, that language will always be merely a part of our everyday language [7, p. 59].

1.3.1. Question of Priority

The purpose of this subsection is to shortly present the question of priority of proving the theorem on undefinability of truth between Tarski and Gödel. This question arose from the fact that Tarski's and Gödel's theorems were consequences of the Diagonalization lemma and as such are closely interrelated [11, p.

153]. For example, that the provable sentences are definable in the first order language of arithmetic, but the truths are not is a consequence of Tarski's theorem. But, the fact that the axioms can be true and the inference rules preserve truth, while at the same time some truths are not provable, is the central point of Gödel's first incompleteness theorem.

Tarski did use Gödel's methods in proving the theorem on the undefinability of truth, but claimed that his results were obtained independently of Gödel's work. Gödel, on the other side, did realize the discrepancy between formal definability of provability and the formal undefinability of truth which ended in discovery of incompleteness [11, p. 157]. A possible explanation for him not explicitly stating the undefinability result can be found in the wide scepticism with regards to semantic notion that was mentioned in Section 1., although when it comes to Tarski, it was precisely that scepticism which motivated him to engage in an examination of the concept of truth. Fur- thermore, Tarski developed a truth-definition and examined the definability of the concept of truth in Peanno's Arithmetic (PA) and stronger theories. On the other side, Gödel had no such truth-definition [25, p. 9] and only considered the problem of expressibility of the set of 'true' numbers in PA.

Woleński [24, p. 459] stresses the importance of using non-finitary devices, as opposed to finitary ones. As Gödel worked on the Hilbert program and his proof of completeness theorem was an important step on the way to finitary metamathematics, while Tarski's semantic definition of truth became the first step toward model theory, whose formulation needed non-finitary devices. Woleński claims that this non-finitary character of Tarski's work is what separates it from that of Gödel and as such only Tarski can claim the discovering of the theorem.

2. The Liar Paradox

The Liar paradox is an umbrella term for a wide range of semantic paradoxes, but in this paper by the Liar paradox we mean those paradoxes which occur because of the explicit self-reference, that some sentences exhibit, resulting in a contradiction

and which in classical logic imply absurdity. That is why the Liar sentences have become the core of arguments against classical logic, since it is because of some key features of classical logic that allow for the existence of the paradox that result in absurdity, through the explosion of consequences [1]. Some of these reactions are the arguments for logics that are paracomplete, which we will cover in Section 4. when talking about Kripke and his theory of truth, and paraconsistent, on which we will give a short introduction in Section 5.

The Liar is constructed by using a truth predicate T, where the predicate T is a truth predicate for language L only if T(⌈ψ⌉) is well-formed for every sentence ψ of L; the conditionals from the Convention T, ψ implies ⌈ψ⌉ and ⌈ψ⌉ implies ψ; and the laws of classical logic. Here we can see the difference between paradoxes and contradictions. Unlike contradictions, which are inconsistent propositions that imply falsity of atleast one of our assumptions, paradoxes are arguments with acceptable and often intuitive assumptions which we deem correct, but which end up in a contradiction [20, p. 198]. That is why paradoxes invite us to reject our previous assumptions; definability of T in the language where we use it in case of Tarski or logical laws in case of paracomplete and paraconsistent logics. Consider the sentence:

4) This sentence is not true.

If 4) is true, then it says of itself that it is not true; and if it is not true, then it says of itself that it is true. Since, in the context of formal languages, every sentence needs to be true or not true, the sentence 4) is both true and untrue, which is a contradiction because it violates the law of non-contradiction.

Formally written, this is how one gets to a contradiction:

1. T(⌈ψ⌉) ∨ ¬T(⌈ψ⌉).

2. Suppose T(⌈ψ⌉).

3. ψ, since ⌈ψ⌉ implies ψ.

4. ¬T(⌈ψ⌉), since ψ says of itself that it is not true.

5. Therefore, T(⌈ψ⌉) ∧ ¬T(⌈ψ⌉).

6. Suppose ¬T(⌈ψ⌉).

7. ψ, since ψ says of itself that it is true.

8. T(ψ), since ψ implies ⌈ψ⌉.

9. Therefore, ¬T(⌈ψ⌉) ∧ T(⌈ψ⌉).

10. Finally, by the disjunction principle we conclude T(⌈ψ⌉) ∧ ¬T(⌈ψ⌉).

Tarski thought that the Liar shows the ordinary notion of truth to be incoherent or too vague, and that it is needed to replace it with a more sci- entifically respectable one. When it comes to formalized languages, Tarski's way of dealing with Liar-type sentences is through a differentiation between metalanguage and object language as we have shown in 2.2. So, because of the presence of the truth predicate T in the object language a Liar sentence is not a well formed formula, which is a result of Tarski's theorem. But, as we will show in 3.1, this is a pyrrhic victory for Tarski.

2.1 Revenge of the Liar

5) This statement is true.

Following Tarski's result, since this sentence contains its true predicate it is not true nor is it not not true. In fact, it is not well formed, since that sentence is not something which we can call true or not true, analogously to a sentence "Tomorrow is (not) true.". Therefore, we can divide sentences on true, false and not well formed and none of the sentences can be in more than one category. So, in which category should the next sentence go:

6) This sentence is not true or it is not well formed.

Since this sentence contains its truth predicate in the same way as a sentence 5), it is not well formed and as such cannot be

true or false. But, it says of itself that it is not well formed, which means that it is true and it can be true only if it is well formed, so we arrive at a contradiction [21, p. 539].

Unlike the cases of sentence 5) where one can say that it is neither true nor not true since it is not a well formed formula, which is tenable, the same approach in this scenario would require that we accept that some sentences are not well formed nor are they well formed. But even if we accepted that, a further problem emerges with a sentence:

7) This statement is not true, or it is not well formed, or it is not well formed nor not well formed.

It soon becomes obvious that by postulating these new differentiations, one falls in to an infinite regress of ever new forms of Liar-type sentences and that even the metalanguage - object language distinction cannot resolve or dissolve the paradox. These revenge problems are common for any proposed solution for Liar-type sentences [1]. Once a solution has been proposed for the Liar paradox, we define Liar sentences with a certain notion, in this case well-formedness. At this point, we use that same notion to create new sentences which we cannot solve since they contain that same notion [2, p. 4]. We shall see in 4.3 and 5.1, how one can construct revenge sentences for other solutions to the "original Liar".

3. Kripke's Theory of Truth

When considering ways to define truth and to find a solution for the Liar paradox, one can turn to the ingredients needed to create the paradox to see what one can change or omit from his system to solve or dissolve the problem. As it was mentioned in Section 3., there is a so called paracomplete approach to this issue. The idea is that the Liar paradox arises from a mistaken intuition formulated in the Law of the Excluded middle (LEM). One such approach was put forward by Saul Kripke and in the core of his proposed theory is the idea that sentences can have a third value, along side true and false, *undefined*.

This theory finds its justification in a critique of the hierarchy of languages developed by Tarski as a result of his Undefinability theorem and Russell, which resulted in his type-theory. By using a three-valued logic, Kripke is also circumventing the consequences of Tarski's theorem, as it was developed in a bivalent system.

Kripke starts his seminal 1975. paper "Outline of a Theory of Truth" by showing that the paradoxality of a sentence needs not to be a matter of its intrinsic, syntactical feature. Kripke invites us to consider the following example [9, p. 691]:

8) Most (i.e., a majority) of Nixon's assertions about Watergate are false.

Intuitively, there is nothing wrong with this sentence in a sense that it has some features because of which we would doubt its well-formedness for example. We can find out the truth value of 8) by collecting all of Nixon's assertions about Watergate and assessing each for truth or falsity. If most of them are false, then 8) is true.

But, Kripke goes on and invites us to suppose that 8) is Jones's sole assertion about Watergate and that the amount of Nixon's true and false assertions about Watergate are evenly balanced, with one exception:

9) Everything Jones says about Watergate is true.

In this case, 8) and 9) are both paradoxical, although we are not dealing with dubious sentences at all. Thus, Kripke concludes, we should not examine the intrinsic qualities that will enable us to pinpoint non-well-formed sentences which lead to the liar paradox [9, p. 692]. It is the empirical fact which creates the paradox and for Kripke, it's not an uncommon thing:

> The versions of the Liar paradox which use empirical predicates already point up one major aspect of the problem: many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features [9, p. 691].

Kripke's proposed theory is based on an analysis of the formentioned problem as well as critique of Tarski's hierarchical solution, which is covered in the next subsection. Although Kripke attacks the idea that each sentence has a fixed level in the hierarchy of languages according to its syntactic form, he still keeps the idea of having a hierarchy but it is not an explicit part of the syntax.

3.1. Critique of Tarski's Hierarchy of Languages

Kripke is presenting a number of objections to the idea of the hierarchy of languages that are mostly grounded in common sense and natural languages. Therefore, when talking about Tarski's Hierarchy of Languages, Kripke is actually talking about a transference of Tarski's object language - metalanguage distinction from formal languages to natural ones and it is worth stressing again that Tarski explicitly states in his 1933. paper that he will not deal with natural languages since he considered them inconsistent by their nature.

The first objection that Kripke presents is the one we mentioned in 4.1. It says that Tarski-type hierarchies assume that the level in the hierarchy which a sentence occupies and the fact of its paradoxality is determined by its syntax, which it is not following Kripke's example with sentences 8) and 9). Therefore, whatever truth theory we are constructing it has to allow statements about truthfulness to be "risky" [9, p. 692] and by that Kripke means that sentences involving truth have a possibility to become paradoxical if the empirical facts create a situation where such thing occurs.

The second objection concerns our ordinary notion of truth, which, commonsensically, should not be an object of stratification, but one, unified concept of 'true'. Further on, Kripke states that in case of the hierarchical approach, one should attach an explicit or implicit subscript to his utterance of true or false, so we can see at which level a sentence is in discourse [9, p. 695]. This feature is not present in our language usage, so it cannot be true of our language. Although this is not a knockout argument against Tarski- type hierarchy, a problem arises when we have two

sentences, each saying something about the truthfulness of the other. Consider Kripke's example:

> Dean says:
> 10) Everything Dean says about Watergate is false. and
> Nixon says:
> 11) Everything Dean says about Watergate is false.

Since 10) says something about the truthfulness of 11) it should be of a higher level than 11) and vice versa. If 10) and 11) are sentences on the same level, neither can talk about the truthfulness of the other, otherwise the higher can talk about the lower, but not conversely. In the same time, there is intuitively no problem with assigning value to these two sentences.

Lastly, Kripke points out that in everyday language usage it is not just the case that we are not using such hierarchies, but it is also not clear how would we use it for epistemological reasons. It is the problem of knowing on which level the assertions of others are, that creates the problem for hierarchies.

3.2. Kripke's Many-valued Solution

Following the issues that arose with 8), Kripke claims that 8) is meaningful, but it can and cannot "express a proposition" [9, p. 700], so we need a semantical scheme for partially defined predicates and a usage of three-valued logic; Kleene's strong three-valued logic is used in Kripke's 1975. paper. In Kleene's strong three-valued logic ¬P is true (false) if P is false (true), and undefined if P is undefined. A disjunction is true if at least one disjunct is true regardless of whether the other disjunct is true, false, or undefined; it is false if both disjuncts are false; undefined, otherwise.

So, in Kripke's theory the truth predicate applies to some, but not all of the sentences of the language and as such the partially defined predicate has truth values of true or false, when it is applied to one of the sentences for which the predicate has been defined, and otherwise it gets the value undefined that is denoted by 'u'. Furthermore, we introduce the notion of partial models [9, p. 700] to accommodate for partially defined

predicates. In a partial model, given a non-empty domain D, an n-place predicate Pn is interpreted by a pair (S_1, S_2) of disjoint n-place relations on the domain of the model (disjoint subsets of D). S_1 is called the extension of P^n and S_2 its anti-extension. In the model, Pn is true of the objects in S_1, false of the objects in S_2, and undefined otherwise. The extension and anti-extension are mutually exclusive, but they do not necessarily exhaust the domain and it is this feature that explains the existence of paradoxical sentences, by putting them outside of the union of extension and anti-extension, $S_1 \cup S_2$.

Kripke constructs a hierarchy of languages of his own, but the level in a hierarchy which a sentence occupies is not determined by its syntax [9, p. 703]. Kripke starts with a language L(Λ,Λ), where Λ denotes an empty set, that is the extension and anti-extension of the truth predicate T are empty sets, so the predicate T is undefined. Then, we build up a hierarchy by defining T with a pair (S_1, S_2) in a way that for any α, the language Lα+1 is like Lα except that T is interpreted by the (S_1, S_2), where S1 is the set of Gödel numbers (ψ) of sentences ψ true in L_α and S_2 is the set of Gödel numbers (ψ) of sentences ψ false in Lα. While building a higher-level language $L\alpha+1$ the interpretation of T is extended by giving it a definite truth value for cases that were undefined in L_α, but no truth value that was defined in Lα can become undefined in $L_{\alpha+1}$. So, the interpretation of T in Lα+1 extends the interpretation of T in L_α and we define a new level of language L_β by T being the union of all the extensions of T in L_0, ... , L_α, where $\alpha \leq \beta$ and the same goes for the anti-extension.

During the construction of these levels, we reach a certain point where $L_\alpha = L_{\alpha+1}$, which we call a fixed point and denote it as L_σ [9, p. 705]. Since it is a fixed point, L_σ is a language that contains its own predicate T, because it assigns T to all of the sentences T can be assigned to and L_σ cannot be further extended. As such, it is an adequate concept for our notion of truth. At this point, some sentence ψ is true in Lσ, if and only if T (ψ) is true in L_σ, which is analogical to the Convention T. This result contradicts the undefinability theorem, because in the same time this construction satisfies Convention T, but still there is a language that contains its predicate T.

Now, Kripke introduces the notion of (un)groundedness. For a sentence ψ of L, ψ is grounded if it has a truth value in L_σ and if ψ does not get a truth value in L_σ it is ungrounded [9, p. 706]. So, the idea is that if some sentence ψ contains a predicate T, its truth value has to be determined by examining the sentences of higher order. If at L_σ we have sentences that do not contain a predicate T and the truth value of ψ can be determined through them, then ψ is grounded, and if not then it is ungrounded [9, p. 694].

Kripke's solution for the Liar paradox is showing that those kind of sen- tences are ungrounded and as such they do not have a truth value. In other words, Liar sentences are undefined or they suffer from a truth-value gap.

3.3. Revenge of the Liar II

In Section 4. a new proposal for solving the Liar paradox was introduced. It says that the Liar-type sentences are undefined and as such they are neither true nor are they not-true. But, as Kripke himself points out in his 1975. paper, the proposed theory suffers from a strengthened version of the Liar in a similar way Tarski's solution does [9, p. 714]. Consider a sentence:

12) This sentence is either false or undefined.

If this sentence is true then it is false or undefined, while if it is false or undefined, then it is true. Hence, 12) is paradoxical. Further on, Kripke defines paradoxical sentences as those that are neither true nor false in a fixed point. Consider a sentence that states just that:

13) The sentence 4) is undefined.

This sentence says something about 4) which is true in Kripke's paracomplete approach. But, during the construction of Kripke's hierarchy, the liar sentences such as 4) never get a truth value and we assign to them the value of undefined only once we have finished the construction. At this point, we cannot add new sentences to the construction. Hence, to express 13) we need to add a truth value of true to it, but we cannot since, 4) becomes un-

defined only after we have already finished our construction [19, p. 221]. We know that 13) is true for L_σ, because in L_σ a Liar sentence is ungrounded and as such it is undefined, but 13) is not expressible in L_σ. Therefore, the paracomplete approach cannot state its own solution to the Liar paradox. In other words, there are truths of L_σ, not expressible in L_σ. This is obvious once we understand that the third value of undefined that we have added is a result of the reflection in the metalanguage on the inability to ascribe one of two classical values to a sentence [5, p. 347]. That means that, in order to express 13), we need to use a metalanguage higher than L_σ analogously to tarskian hierarchy [9, p. 714].

4. Paraconsistent Logic and Truth

Until now, we have examined two different approaches to truth and semantic paradoxes. One, here presented by Tarski, is a classical, or orthodox, approach by which we mean obeying the rules of classical logic, most notably its bivalence. The second one, presented through Kripke, is a paracomplete one, which means that we are dropping the law of the excluded middle in a sense of adding a third value along side true and false, that of undefined, to accommodate the fact that a sentence can be neither true nor false. The last that we will cover in this paper is the paraconsistent approach.

In paraconsistent logic it is the principle of explosion of consequences that is rejected. So, $A, \neg A \vdash B$ does not hold. In other words, contradictions do not imply absurdity [14, p. 75]. It is important to note that a certain logic can be both paraconsistent and paracomplete, but in this section we will concentrate on those that are paraconsistent and non-paracomplete. One such logic is proposed and defended by Priest in his 1979. paper "The Logic of Paradox" [16], where he devises such a system and calls it Logic of Paradox; LP.

Although LP can be understood as a three-valued logic, analogously with Kleen's strong three-valued logic Kripke used in his 1975. paper, in case of LP there are no truth-value gaps, but truth-value gluts [14, p. 109]. That means that, instead of the third

value being neither true nor false, the value is both true and false. Still, there is a substantial difference between the two three-valued logics, moreover Priest even criticises the truth-value gap approach.

The critique runs as follows, if there is no fact that makes a sentence ψ true, then there is a fact that makes $\neg\psi$ true, and that fact is that there is no fact that makes ψ true [15, p. 65]. In other words, suppose that ψ is a sentence, and suppose that there is nothing in virtue of which ψ is true. Then this is something in virtue of which $\neg\psi$ is true, even though we may not know what it is. From this it follows that a sentence ψ has to have a truth value of true or false, therefore there are no truth-value gaps. But this does not mean that ψ cannot have both truth values, or have a truth- value glut. In other words, accepting truth-value gluts means accepting the existence of true contradictions. A position which defends such a view is called *dialetheism*.

Paraconsistent logics can produce their own interpretation for the truth predicate in a similar manner as Kripke did with his hierarchy of interpretations [1]. The idea is to start with a language L_0, which does not contain its truth predicate and the extension and anti-extension of the truth predicate T are empty sets. We then build up our hierarchy by extending the exten- sion and anti-extension with new sentences. The main difference between this construction and that of Kripke is that the extension and anti-extension are exhaustive but not mutually exclusive. So, there is no such thing that lies outside the extension and anti-extension, but there are things that are members of both. This is the idea of a glut mentioned before.

There are a number of strengths that this kind of approach exhibits. First of all, we have seen that Tarski's Convention T easily leads to contradictions, which are blocked by Tarski by saying that the predicate T cannot be ex- pressed in the object language, so a liar sentences cannot be formulated. But, one can construct a paraconsistent Tarskian truth-theory which contains its own predicate T [14, p. 46]. Second, while other theories are constructed so they would solve or dissolve the paradoxical sentences, those same sentences are used as arguments for paraconsistancy and dialetheism [16, p. 220]. Com- pared to Kripke's paracomplete approach, while Kripke cannot state its

own solution to the paradox, the paraconsistent one seems to not have any issues with it. Thirdly, if logic wants to examine natural reasoning in natural languages, its models should coincide with those natural phenomenons. Tarski claimed that the universality of everyday language is the source of seman- tic antinomies, like that of the liar. Therefore, if we endorse such a view of natural languages, then we should also endorse a dialetheist position since it accommodates how natural language works [16, p. 225].

4.1. Revenge of the Liar III

Even though the paraconsistent, or dialetheist, solution seems elegant, although definitely not non-contentious, it is not completely free from some issues. Revenge of the Liar is an adequacy objection in a sense that a lan- guage is inadequate due to its expressive limitation and in the similar way as with Tarski's and Kripke's solution, we have an expressivity problem with regards to dialetheist position [3, p. 743].

The position presented in this Section holds the following true:

14) Every sentence is either only true or only false or both true and false.

But, as Bromand showed [3, p. 745], Priest's account, like those of Tarski and Kripke, introduces novel kinds of semantic concepts which cannot be adequately expressed in that same approach. In this case, it is the notions of *only true* and *only false* that are contentious. The problem lies in the fact that, when these notions are expressed in the language in question, we are led to triviality. In other words, since only true and only false cannot be properly expressed, as a result we get that being only true is equal to only false, that is, every sentence becomes true. For example, lets show the problem with *only false*. A sentence ψ is *only false* if and only if ψ is not true in L. Assume that L contains a predicate Fx that defines ψ as ψ is *only false* in L. There will be a sentence ψ that says F($\lceil\psi\rceil$). We can prove in the metalanguage that ψ is true in L if and only if F($\lceil\psi\rceil$) is true in L if and only if ψ is *only false* in L

if and only if ψ is not true in L. But, sentences of L cannot be both true and not true in L. Therefore, L cannot express *only false* in L.

5. Pursuit of Truth

In this paper, we examined three approaches to the notion of truth, as well as three proposals for solving Liar-type sentences. Now, in this last section, we want to take a step back and look at what these approaches can tell us about the link between logic, language and truth.

As we have seen in previous sections, for accomplishing the goal of defining truth we are constructing a formal account of truth by building models.

The purpose of these models is to capture certain features of natural lan- guages and of our ordinary notion of truth, like those of capture ($\psi \vdash T(\lceil\psi\rceil)$ and release ($T(\lceil\psi\rceil) \vdash \psi$), that are central for constructing Liar-type sentences for example, or intersubstitutability of T (ψ) and ψ, that the attempts of internalizing the concept of truth we examined wanted to achieve and maintain. So, these models are idealizations of natural languages, while the formalizations of truth are about capturing our intuitions of truth in the truth predicates of those models. Now, the problem is in detecting, if these models function in a similar enough way as natural languages and if the truth predicates function in a similar enough way as truth does in natural languages. Note that we say in a similar enough way because we are talking about idealizations of natural languages. The problem of detecting this compatibility of features is actually the problem of adequacy. Through this paper, we have examined the problem of adequacy through Liar-type sentences.

We have shown why and how both natural and formal languages give rise to Liar-type sentences. Their existence in undoubted and what an adequate model needs to show is how, despite these paradoxical sentences, our truth predicate achieves the features we want it to have and what are the relevant features of our languages that are involved in it. The issue that repeatedly reemerged is the revenge problem. No matter which approach to Liar-type sentences we examined, it turned out that the success of

a particular solution was due to the fact that the model language we were dealing with achieved the features we wanted it to have in virtue of lacking the expressive power of natural languages. Since languages we built lack some important features of natural languages, those models are inadequate.

Three approaches to these problems that were covered in this paper are orthodox, paracomplete and paraconsistent. The orthodox approach was represented by Tarski and its name comes from the fact that it maintains the rules of classical logic. As it was demonstrated by Tarski's theorem, the truth predicate cannot be defined in the object language, so a hierarchy of fixed levels is created to accommodate for the use of the notion of truth. Thus, Liar-type sentences, since they contain a truth predicate and talk about themselves, are not well-formed and the Liar paradox is dissolved. Kripke's paracomplete approach differs by rejecting the law of the excluded middle, that is by rejecting bivalence. Alongside true and false, there is a third value of undefined that is ascribed to sentences which are ungrounded, for example the Liar-type sentences. The third approach is paraconsistent, here represented by Priest. Paraconsistency means rejecting the explosion principle, or in other words not everything is provable from a contradiction. Although it is not necessarily so, paraconsistent approaches can be viewed as having three values, like in the case of paracomplete logic, but with the third value being both true and false, again an example being Liar-type sentences.

What all of these approaches have in common is that they are inade- quate in a sense as used above. There are some important features of natural languages that these models do not achieve. The feature that was shown in this paper is expressivity, or more precise being able to express its own result. This was demonstrated by producing strengthened Liar sentences, also known as revenge of the Liar. Expressing these sentences required from each language to use semantic notions they introduced to solve the Liar paradox in a new setting, which resulted in new paradoxes. Nevertheless, the work done was not in vein. As it was explained in the Introduction, the notion of truth has transformed from a forbidden term to a concept that is widely discussed and used in a range of philosophical disciplines.

References

1. Beall, J., Glanzberg, M., Ripley , D. (2016) Liar Paradox, The Stan- ford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = https://https://plato.stanford.edu/entries/liar- paradox/.
2. Beall, J. (2007) "Prolegomenon to Future Revenge", In: Revenge of the Liar - New Essays on the Paradox, ed. J. C. Beall, New York: Oxford University Press, p. 1–31.
3. Bromand, J. (2002) "Why Paraconsistent Logic Can Tell Only Half of the Truth", Mind 111, 741–749.
4. Carnap, R. (1942) Introduction to Semantics, Cambridge: Harvard Uni- versity Press, p. VIII.
5. Čulina, B. (2001) "The Concept of Truth", Synthese 126, 339–360.
6. Heck, R. G. (1997) "Tarski, Truth, and Semantics", The Philosophical Review 106, 533–554.
7. Kovač, S. (2005) Logičko-filozofijski ogledi, Zagreb: Hrvatsko filozofsko društvo.
8. Kovač, S. (2013) Svojstva klasične logike, Zagreb: University Centre for Croatian Studies, University of Zagreb, Manualia, Vol. 10.
9. Kripke, S. (1975) "Outline of a Theory of Truth", The Journal of Phi- losophy 72, 690–716.
10. Macan, I. (1997) Filozofija spoznaje, Zagreb: Filozofsko-teološki institut Družbe Isusove.
11. Murawski, R. (1998) "Undefinability of Truth. The Problem of the Pri- ority: Tarski vs. Gödel", History and Philosophy of Logic 19, 153–160.
12. Murawski, R. (2006) "Troubles with the Concept of Truth in Mathe- matics", Logic and Logical Philosophy 15, 285–303.
13. Niniluoto, I. (1999) "Theories of Truth: Vienna, Berlin, and Warsaw". In: Alfred Tarski and the Vienna Circle: Austro-Polish Connections in Logical Empiricism, eds. J. Wolenski, E. Köhler, Vienna: Springer Sci- ence+Business Media Dordrecht, p. 17–26.
14. Priest, G. (2005) Doubt Truth to be a Liar, New York: Oxford University Press.
15. Priest, G. (2006) In Contradiction: A Study of the

Transconsistent, New York: Oxford University Press.
16. Priest, G. (1979) "The Logic of Paradox", Journal of Philosophical Logic
8, 219–241.
17. Soames, S. (2010) Philosophy of Language, Princeton: Princeton Uni- versity Press.
18. Soames, S. (1999) Understanding Truth, Oxford: Oxford University Press.
19. Storm Hansen, C. (2014) "Grounded Ungroundedness", Inquiry An In- terdisciplinary Journal of Philosophy 57, 216–243.
20. Šikic, Z. (1992) "The Diagonal Argument - a Study of Cases", Interna- tional Studies in the Philosophy of Science 6, 191–203.
21. Švob, G. (1985) "Ima li danas logičkih antinomija?", Filozofska istraži- vanja 3, 527–542.
22. Tarski, A. (1956) "The concept of truth in formalized languages", In: Logic, Semantics, Metamathematics, ed. A. Tarski, Oxford: Oxford Uni- versity Press, p. 152–278.
23. Tarski, A. (1983) The Establishment of Scientific Semantics, In: Logic, Semantics, Metamathematics, ed. A. Tarski, Oxford: Oxford University Press, p. 401–409.
24. Woleński, J. (2005) "Gödel, Tarski and Truth", Revue internationale de philosophie 4, 459–490.
25. Woleński, J. (1999) Semantic Revolution - Rudolf Carnap, Kurt Gödel, Alfred Tarski, In: Alfred Tarski and the Vienna Circle: Austro-Polish Connections in Logical Empiricism, eds. J. Wolenski, E. Köhler, Vienna: Springer Science+Business Media Dordrecht, p. 1–16.