# Kripke's Theory of Truth and the Liar Paradox

Doesik Kim (Konkuk University)

【Abstract】
The purpose of this paper is to defend Kripke's theory of truth from Simmons' objection. First, after introducing various sorts of the liar paradox, briefly I explain Tarski's attempt to solve the puzzle. Then, I outline Kripke's solution by using the concept of 'fixed point'.

Simmons offers an interesting objection against Kripke's solution. He uses a diagonal argument in his attack to Kripke's idea. I claim that Simmons seem to use 'exclusion negation' in refuting Kripke. I think, however, there is an alternative interpretation, which is 'choice negation'. With using choice negation, I maintain that Kripke's theory of truth can be defended from Simmons' objection.

【Key Words】
Kripke, Theory of Truth, the Liar Paradox, Diagonal Argument

【국문 제목】
크립키의 진리론과 거짓말쟁이의 역설

## I.

The Liar sentence is the sentence which leads to contradiction from apparently obvious principles about truth, by an apparently valid inference. This is the reason a paradox occurs. I guess the name 'Liar's paradox' comes from the following example which is also called 'the Epimenides paradox'. Suppose a Cretan whose name was Epimenides said, "All Cretans are liars". If what Epimenides said is true, then he is also a liar, so what he said is false. But this is not completely paradoxical since it can be consistently supposed to be false. Namely, if it is false, what he said means there is at least one Cretan who is not a liar: this does not lead to a contradiction.

The classical version of the Liar's paradox is this:

(S) This sentence is false.

If S is true, then what it says is the case: so it is false. On the other hand, if S is false, then what it says is not the case, so it is true. In other words, S is true iff S is false, but this is an apparent paradox.

Let me introduce another version of Liar which involves a 'semantic' concept.[52] Let's say 'heterological' means 'not true of itself'. 'Long', 'monosyllabic', 'German' are heterological while 'short', 'English' are autological, namely true of themselves. But if 'heterological' is heterological, it is not true of itself, so it is not heterological. On the other hand, if 'heterological' is not heterological, it is true of itself, so it is heterological. Now we come to a contradiction again.

In order to solve these paradoxes, two requirements are needed. First, the solution should give a consistent formal theory of semantics, namely it should indicate which apparently unexceptionable premises or principle of inference must be given up. Second, it should supply some explanation of why that premise or principle is exceptionable in spite of its appearance.[53]

The most popular solution on Liar's paradox is made by Tarski. He thinks the semantic paradoxes come from the following two assumptions:

a) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term 'true', referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "semantically closed".

b) We have assumed that in this language the ordinary laws of logic hol d.[54]

52) There are other paradoxes related to the set theoretical concept such as Russell's paradox, Cantor's paradox, Burali-Forti paradox, etc. which are beyond the scope of this paper.

53) S. Haack, Philosophy of Logics, 1978, pp. 138-139.

Tarski denies that an adequate definition of truth can be given for a semantically closed language, as he proposes a formal adequacy condition, which says truth can be defined only for semantically open languages. So Tarski proposes a hierarchy of languages: the object language O which is "talked about" and the meta-language in which we "talk about" the object language. The meta-language contains means of referring to expressions of the object language and the predicate 'true-in-O' and 'false-in-O'. It should be noticed that these terms 'object language' and 'meta-language' have only a relative sense, in which we can construct a whole hierarchy of language s.[55]

Since truth for a given level is always expressed by a predicate of the next level, the Liar sentence such as 'This sentence is false' can appear only in the harmless form 'This sentence is false-in-O'. If so, 'This sentence is false-in-O' must be a sentence of meta-language, which cannot be true-in-O, therefore it is simply false instead of paradoxical.

Though Tarskian approach seems to be accepted by many philosophers, there are some criticisms about its artificiality. In other words, the Tarskian explanation of Liar sentences only gives a formal solution but not a philosophical one, because the reason Tarski gives for requiring semantic openness is simply that the semantic closedness leads to contradiction. Thus Tarskian solution does not seem to be quite suitable for explaining our natural language which has truth predicate in itself. And Tarskian assertion that there are different 'true' predicates at each level does not correspond to our intuition: we have only one 'true' predicate. This is why Kripke is not satisfied by Tarskian approach.

---

54) A. Tarski, "The Semantic Conception of Truth" in Philosophy and Phenomenological Research 4, 1944, pp. 348-349.

55) ibid, pp. 349-351.

# II.

Kripke does not think of 'true' as systematically ambiguous in the way Tarskian solution suggested. According to Kripke, in his paper "Outline of a Theory of Truth", ordinary ascriptions of truth and falsity cannot even be assigned implicit levels. And if the empirical facts are extremely unfavorable, Kripke argues, many of our ordinary assertions are apt to turn out paradoxical. Suppose, for instance, that John says:

(1) Most of Nixon's assertions about Watergate are false.

The truth value of this assertion should be assigned to the next level above the highest level of any of Nixon's utterances about Watergate. But not only do we have no way to determine the levels of Nixon's utterances about Watergate in an ordinary sense, but also it may be impossible to assign levels consistently. Suppose that (2) is one of Nixon's utterances about Watergate:

(2) Everything John says about Watergate is true.

Then John's utterance, say (1), has to be at a level one higher than all of Nixon's, and Nixon's utterance, say (2), should be in one higher level than all of John's. Furthermore, if we are allowed to say that Nixon's assertions about Watergate are evenly balanced between the true and the false except (2), and in addition, to say that all John's assertions related to Watergate are true except (1), then both (1) and (2) are paradoxical: they are true iff they are false.[56] Kripke asserts that Tarskian approach fails to take adequate account of the 'risky' character of truth. Even if, clearly, nothing is

56) S. Kripke, "Outline of a Theory of Truth", The Journal of Philosophy, 1975. Reprinted in Truth and the Liar, ed. by R. L. Martin, 1984, pp 54 - 55 and pp. 59 - 60.

intrinsically wrong with (1), nor is it ill-formed, it can be a paradoxical assertion under unfavorable empirical circumstances.

Another problem with Tarskian approach, Kripke says, is that of transfinite levels. With Tarskian solution, it is easy to assert that snow is white, 'Snow is white' is true, "'Snow is white' is true" is true, etc.: the various occurrences of 'is true' in the sequence are assigned increasing subscripts. But it is very hard to claim that all the statements in the sequence are true. In order to do so, we need to have a metalanguage of transfinite level, above all the languages of finite level. But Kripke thinks Tarski's hierarchy of languages is only for finite levels, which is inadequat e.[57]

Kripke seeks to supply an explanation of the source of paradox which is more satisfactory in this respect, and then to build up a formal theory on this basis. His idea is to deny that the truth predicate must be totally defined, that is to say, that every well-formed sentences must be either true or false. The key idea in this explanation is the concept of groundedness. Kripke explains his idea as follows:

> If a sentence such as (1) asserts that (all, some, most, etc.) of the sentences of a certain class C are true, its truth value can be ascertained if the truth values of the sentences in the class C are ascertained. If some of these sentences themselves involve the notion of truth, their truth value in turn must be ascertained by looking at other sentences, and so on. If ultimately this process terminates in sentences not mentioning the concept of truth, so that the truth value of the original statement can be ascertained, we call the original sentence grounded, otherwise, ungrounded. As the example of (1) indicates, whether a sentence is grounded is not in general an intrinsic (syntactic

57) ibid, pp. 60 - 61.

or semantic) property of a sentence, but usually depends on the empirical facts.[58)]

Kripke tries to give, by accepting truth value gaps, an interpretation of the ordinary use of 'true' predicate and a solution to the Liar paradox. Kripke shares the common feature with Tarski in the respect of accepting a hierarchy of interpreted languages where, at any level, the truth predicate is the truth predicate for the next lowest level. At the lowest level the predicate 'true' is completely undefined. At the next level, it is assigned to true well-formed formulas which do not themselves contain 'true'. But Kripke's idea is different from that of Tarski's in the following sense. According to Kripke, there is a fixed point at which a language contains its own truth predicate. He seems to think that this 'true' predicate is a natural choice for modelling 'true' in ordinary language.

I will explain how Kripke expresses his ideas in formal expression. Let L be an interpreted, classical first-order language with a finite (or even denumerable) list of primitive predicates. We assume that the variables range over some non-empty domain D, and the primitive n-ary predicates are interpreted by (totally defined) n-ary relations on D.[59)]

Now we extend L to a language $L^*$ by adding a monadic predicate $T(x)$ whose interpretation need only be partially defined. An interpretation $T(x)$ is given by a "partial set" $(S_1, S_2)$ where $S_1$ is the extension of $T(x)$, $S_2$ is the anti-extension of $T(x)$, and $T(x)$ is undefined for entities outside $S_1 \cup S_2$. Let $L^*(S_1, S_2)$ be the interpretation of $L^*$ which results from interpreting $T(x)$ by the pair $(S_1, S_2)$, the interpretation of other predicates of L remaining as before. Let $S_1'$ be the set of true sentences of $L^*(S_1, S_2)$ and let $S_2'$ be the set of all elements of D which either are not sentences of $L^*(S_1, S_2)$ or are false sentences of $L^*(S_1, S_2)$. $S_1'$ and $S_2'$ are uniquely

58) S. Kripke, ibid, p. 57.
59) ibid, p. 66.

determined by the choice of $(S_1, S_2)$. Clearly, if T(x) is to be interpreted as truth for the very language L containing T(x) itself, we must have $S_1 = S_1'$ and $S_2 = S_2'$. This is what Kripke called a 'fixed point'. For a given choice of $(S_1, S_2)$ to interpret T(x), set $\varnothing((S_1, S_2)) = (S_1', S_2')$. $\varnothing$ then is a unary function defined on all pairs $(S_1, S_2)$ of disjoint subsets of D, and the 'fixed points' $(S_1, S_2)$ are literally the fixed points of $\varnothing$. 60)

Here, $\varnothing$ is a monotone operation on $\leq$: that means if the interpretation of T(x) is extended by giving it a definite truth value for cases that were undefined, no truth value previously established change or becomes undefined, namely, at most, certain previously undefined truth values become defined. 61)

Given the monotonocity of $\varnothing$, we can deduce that for each (finite) a, the interpretation of T(x) in $L^*_{a+1}$ extends the interpretation of T(x) in $L_{*a}$. We can do it in transfinite level, too. Formally, we can define the language $L_*a$ for each ordinal a. If a is a successor ordinal (a=b+1), let $L_{*a} = L_*(S_{1,b}, S_{2,b})$. If l is a limit ordinal, $L_{*l} = L_*(S_{1,l}, S_{2,l})$ where $S_{1,l} = U_{b<l}\, S_{1,b}$ and $S_{2,l} = U_{b<l}\, S_{2,b}$. So, at "successor" levels we take the truth predicate over the

---

60) ibid, pp. 66 - 67.

61) Kripke accepts Kleene's strong three-valued logic. There are some other schemes, for example Kleene's weakened valuation scheme. (This is the same as Bochvar's three-valued system so far as truth table is concerned.)

And there is another matrices by Lukasiewicz. In this scheme, ~P, P v Q, P & Q are the same as Kleene's strong three-valued logic. But an interesting thing is the fact that the value of P → Q, when both P and Q are undefined, is true, while in Kleene's system, the value of P → Q is undefined.

If Kripke were to use Lukasiewicz's scheme, he cannot explain 'monotonocity' due to the following fact. Suppose, at some level n, P and Q are both undefined. According to Lukasiewicz's scheme, P → Q is true. But if P is defined as true and Q as false at the next level, n+1, the truth value of P → Q turns out to be false. In this case, the truth value previously established is changed. This contradicts the monotonocity.

This explanation may elucidate some features on monotonocity, namely under what conditions the monotonocity can be maintained, and so on. But that is not the task in this paper. (cf. S. Haack, Philosophy of Logics, 1978, pp. 205-208, & S. C. Kleene, Introduction to Metamathmetics, 1952, pp. 332-340)

One more comment on this: Kripke does not regard his use of Kleene's three-valued valuation rules as a challenge to a classical logic. (See, Kripke's footnote 18, pp 64-65 in "Outline of a Theory of Truth".)

previous level, and at transfinite level, we take the union of all sentences declared true or false at previous levels. Thus even with the transfinite levels included, it remains true that the extension and the anti-extension of T(x) increase with increasing a.

Now we come to the place where we prove that there is a fixed point. Towards a contradiction, suppose there is no fixed point, namely there is no a such that $L_{*a} = L_{*a+1}$. Then, on whatever level a, except base level, there is a sentence which is newly declared true or false at that ordinal level. In this function whose domain is all the sentences which can be declared true or false, and whose range is ordinals at which a sentence can be declared true or false, there should be the set of all ordinals by Axiom of Replacement. [62] But Burali-Forti Theorem says there is no set to which every ordinal number belongs. [63] We come to a contradiction. Therefore, there is a fixed point.[64]

All paradoxical sentences are ungrounded, according to Kripke, but not all ungrounded sentences are paradoxical: a paradoxical sentence is one that cannot consistently be assigned a truth value at any fixed point. This explains why 'This sentence is true' seems to share some of the oddity of 'This sentence is false', and yet, unlike the Liar sentence, can be given a truth value consistently. A truth value can be given to 'This sentence is true', but only arbitrarily, and a truth value cannot be consistently be assigned to 'This sentence is false'.

This picture solves several problems Tarski has. First of all, his idea allows for the 'riskiness' of truth-ascription: for the paradoxical character of a sentence may be either intrinsic as in the case, 'This sentence is false', or empirical as I explained with the sentences, (1) and (2). This is possible for Kripke because he accepts only on truth-predicate rather than truen, for

62) Cf. Enderton, Elements of Set Theory, 1977, p. 179.
63) Cf, ibid, p. 194.
64) Kripke, ibid, pp. 63 - 69.

every level n, as Tarski did. Secondly, his idea also gives an account of transfinite level. If Kripke's explanation is successful, we can have an account of a language which has 'true' predicate in itself, which is closer to our natural language than Tarskian language is.

## III.

In one of the ways of criticizing Kripke's view, diagonal argument plays an important role. Then what is the diagonal argument? The diagonal argument is introduced by Cantor. [65] One of the Cantor's purposes is to show that the reals are nondenumerable. Another purpose is to extend this result to a general theorem, that any set can be replaced by another bigger set using the concept of power set. I will explain diagonal argument with Cantor's first proof.

Consider the two elements m and w (m is different from w). Let M be the set whose elements E are sequence $\langle x_1, x_2, x_3, \ldots\ldots, x_v, \ldots\rangle$ where each of $x_1, x_2, x_3, \ldots\ldots, x_v, \ldots$ is either m or w. Cantor arranged a denumerable list of elements of M in an array:

$$E_1 = \langle a_{1.1}, a_{1.2}, \ldots\ldots, a_{1.v}, \ldots\rangle$$
$$E2 = \langle a_{2.1}, a_{2.2}, \ldots\ldots, a_{2.v}, \ldots\rangle$$
$$E3 = \langle a_{3.1}, a_{3.2}, \ldots\ldots, a_{3.v}, \ldots\rangle$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$En = \langle a_{n.1}, a_{n.2}, \ldots\ldots, a_{n.v}, \ldots\rangle$$

where each $a_{n.v}$, is either m or w. Then we can make an array as follows:

65) G. Cantor, "Über eine elementare Frage der Mannigfaltigkeitslehre", first published in Jahresberecht der Deutschen Mathematicker-Vereinigung 1, 1890-1 pp. 75 - 78 and reprinted in Gesammelte Abhandlungen mathematischen und philosophischen Inhalts, ed. by E. Zermelo, 1932, pp. 278 - 281. I learned Cantor's proof from "The Diagonal Argument and the Liar", Journal of Philosophical Logic, 1990 by K. Simmons, pp. 2 - 9.

| | 1 | 2 | 3 | ....... |
|---|---|---|---|---|
| E1 | m | w | m | ....... |
| E2 | w | m | m | ....... |
| E3 | m | w | w | ....... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

⟨Figure 1⟩ (Here, m or w is arbitrarily assigned.)

This array (R) is composed of two collections - the 'side' (D1) which is the set whose members are $E_1$, $E_2$, $E_3$, .... and the 'top' (D2) which is the set of natural numbers, - and 'values' m and w. There is only one value for any pair of elements taken from the side and the top.

In Cantor's first proof, the diagonal considered is the **leading diagonal**, the cell of which are given by ⟨$E_1$, 1⟩, ⟨$E_2$, 2⟩, ⟨$E_3$, 3⟩, ..... The leading diagonal, from top left to bottom right, corresponds to our intuitive notion of a diagonal, but it is just one of the diagonals. What is important here is a one to one correspondence between $D_1$ and $D_2$. Therefore we define that F is a diagonal on $D_1$ and $D_2$ iff F is a one to one function from $D_1$ into $D_2$ or from $D_2$ into $D_1$.

Now we can distinguish three kinds of diagonal: **complete diagonals, $D_1$-complete diagonals, $D_2$-complete diagonals**. F is a **complete diagonal** on $D_1$ and $D_2$ iff F is a one to one correspondence **between** $D_1$ and $D_2$. (Both Cantor's proofs use complete diagonals.) F is a **$D_1$-complete diagonal** on $D_1$ and $D_2$ iff F is a one to one correspondence **from $D_1$ into $D_2$**. F is a **$D_2$-complete diagonal** on $D_1$ and $D_2$ iff F is a one to one correspondence from **$D_2$ into $D_1$**.

This notion of a diagonal has to do only with position, and yet, there is no link between the cells which constitute the diagonal and the value associated with each cell. The value of the diagonal is, simply speaking, the value each cell has in itself. As an example, in Figure 1, the value of the diagonal is m, m, w, .... . From this value of the diagonal, we can get the countervalue of the diagonal. The procedure is fairly simple. In the above example where we get the value of the diagonal as m, m, w, ....., if the value of the diagonal is m then the countervalue of the diagonal is w, and vice versa. So we can get the countervalue of the diagonal as w, w, m, .... .

Now we arrive at the **diagonal theorem**:

> Let R be an array on $D_1$ and $D_2$ and let F be a $D_1$-complete diagonal on $D_1$ and $D_2$. Let H be a countervalue of F. Then H does not occur as a row of R.

The reason why H cannot occur as a row of R is intuitively simple. In the above example, let the countervalue of the diagonal $E_0$ = ⟨w, w, m, .....⟩. Then $E_0$ cannot occur as a row of R, because $E_0$ is different from $E_1$ in the first digit (since $w \neq m$) and $E_0$ is different from $E_2$ in the second digit (since $w \neq m$) and $E_0$ is different from $E_3$ in the third digit (since $m \neq w$), and generally speaking, $E_0$ is different from $E_n$ (for every natural number n) in the n-th digit. In other words, suppose $E_0 = E_n$ (for some n), then the value of the n-th cell of $E_0$ is identical with the value of the n-th cell of $E_n$, which contradicts the definition of the countervalue. Thus the diagonal theorem is proved. [66]

Before criticizing Kripke's approach by means of the diagonal argument, I will briefly explain what the relation is between the diagonal argument and Liar sentence with the example of 'heterological'. Let the side and the top be the set of one-place predicates of English, of which 'heterological' is a member. The array R can be given by:

$$R(x,y) = \begin{cases} t, \text{ if } x \text{ is true of } y \\ f, \text{ if } x \text{ is not true of } y \end{cases}$$

So the array is:

66) cf. K. Simmons, ibid, p. 10.

| | long | short | monosyllabic | .......... |
|---|---|---|---|---|
| 'long' | f | t | t | .......... |
| 'short' | f | t | t | .......... |
| 'monosyllabic' | t | f | f | .......... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

The values in the table are assigned according to the following rules. The word 'long' is not long. So f is assigned in the cell. And 'long' is short. So t is assigned where the side is 'long' and the top is 'short'. Likewise 'long' is monosyllabic and therefore t is assigned in the cell of the first row and the third column.

The diagonal F here is a kind of identity. A countervalue H of F is given by:

$$H(x,x) = \begin{cases} \text{t, if x is not true of x} \\ \text{f, if x is true of x} \end{cases}$$

which is in this case, t, f, t, ...... . The diagonal theorem says that there is no predicate of English true of exactly those predicates which are not true of themselves. But 'heterological' is such a predicate, so we come to a contradiction. This result is due to the assumption that we can fill the row (and column) of the value associated with the predicate 'heterological'. The diagonal argument shows that the predicate 'heterological' is inexpressible in English without any paradox.[67] This is the same result as we have seen in chapter I.

A way out from this problem is to say that 'heterological' is neither true nor false of itself: the Liar sentence "'heterological' is heterological" has a truth value gap. This is what Kripke did in his paper. Now we come to the stage where Kripke's argument is criticized by the diagonal argument.

67) K. Simmons, ibid, pp. 15 - 16.

# IV.

As I have argued in chapter II, Kripke has shown that, by allowing truth value gaps, we could obtain a formal language which contained its own truth predicate in itself. In other words, the formal language constructed by Kripke can express its own concept of truth without using metalanguage.

But we can point out a problem with any truth value gap approach by means of the diagonal argument. Let's look at the 'superheterological paradox' in three valued array in which we can put a 'u' (undefined) in the 'heterological'/heterological cell. The array is given by:

$$R(x,y) = \begin{cases} t, \text{ if } x \text{ is true of } y \\ f, \text{ if } x \text{ is false of } y \\ u, \text{ if } x \text{ is neither true nor false of } y. \end{cases}$$

So the array is:

| | long | short | heterological | .......... |
|---|---|---|---|---|
| 'long' | f | t | t | .......... |
| 'short' | f | t | f | .......... |
| 'heterological' | t | f | u | .......... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Again, since ⟨'long' is long⟩ is false, ⟨'long' is heterological⟩ is true. On the contrary, ⟨'short' is heterological⟩ is false because ⟨'short' is short⟩ is true. Here, what we have to take a notice is the cell of the third row and the third column.

One of the countervalues is obtained by:

$$H(x,x) = \begin{cases} f, \text{ if } x \text{ is true of } x \\ t, \text{ if } x \text{ is false of } x \\ t, \text{ if } x \text{ is neither true nor false of } x. \end{cases}$$

In this example, the antidiagonal goes t, f, t, ...... . Each t on this antidiagonal is associated with a word which is either false of itself, or neither true nor false of itself. So the associated countervalue corresponds to the concept 'false or neither true nor false of itself'. But the diagonal theorem shows that no English predicate is true of exactly those words 'false or neither true nor false of themselves'. However, there is such a predicate, that is 'is false or undefined of itself'. Therefore we arrived at the superheterological paradox. As a conclusion, the diagonal argument shows that there are several semantic concepts that Kripke's language cannot express. (The reason such semantic concepts are not just one is the fact that there is more than one countervalues.) This implies that Kripke's language is still expressively incomplete with regard to its own semantics, even though Kripke succeeds in constructing a language which has truth predicate in itself. [68] His language is expressively incomplete since Kripke cannot express these semantical concepts without using metalanguage.

It seems to me that there is an exit for Kripke to escape from this predicament. The notion of negation used in the diagonal argument is **exclusion negation**. According to exclusion negation, ~A is true iff A is false or undefined. ~A is false iff A is true. An alternative way of interpreting negation is **choice negation**, which says ~A is true iff A is false, ~A is false iff A is true and ~A is undefined iff A is undefined. Kripke's construction of the minimal fixed point depends upon Kleene's strong three valued scheme which uses choice negation, as I pointed out in footnote 10.

68) K. Simmons, ibid, pp. 21 - 22.

But the diagonal argument used exclusion negation.

If choice negation is accepted, we have no chance to construct the concept 'false of itself or neither true nor false of itself' as an antidiagonal. Why? The main idea in the diagonal argument is to construct an antidiagonal which is different from every element in a row, thus which cannot appear in a row. But when one tries to build a countervalue with truth value gap, he should be very careful. In the previous example in chapter III, the countervalue H can be given at random in so far as it is a countervalue. In other words, there is no further restriction in constructing a countervalue except for what was described in chapter III. Let me explain in details. In heterological paradox, without doubt, it is true that $\sim t = f$ and $\sim f = t$. In Cantors theorem, it is obviously true to say that $\sim n = n+1$ and $\sim 9 = 0$.[69] But the fact that Kripke introduced 'undefined' as a truth value made the story totally different. We assigned countervalue H as:

$$H(x,x) = \begin{cases} f, \text{ if } x \text{ is true of itself} \\ t, \text{ if } x \text{ is false of itself} \\ t, \text{ if } x \text{ is undefined of itself} \end{cases}$$

but we are not allowed to do that. Because, in this case, $\sim u = t$ contradicts Kleene's three valued scheme, according to which $\sim u = u$. In other words, the only way the countervalue H can be given is:

$$H'(x,x) = \begin{cases} f, \text{ if } x \text{ is true of itself} \\ t, \text{ if } x \text{ is false of itself} \\ u, \text{ if } x \text{ is undefined of itself} \end{cases}$$

69) In this case, the countervalue can be given by:

$$H(x,y) = \begin{cases} n+1, \text{ if the number of y-th digit of x is not 9} \\ 0, \text{ if the number of y-th digit of x is 9.} \end{cases}$$

Different from other examples, other ways of constructing a countervalue are blocked by Kleene's scheme in this case. And the only countervalue which is constructed by H'(x,x) may occur in the row. Therefore we do not have any contradiction. Kripke's view is still alive.

Kripke argues that even though the ascent to a metalanguage cannot be avoided, his theory provides a model for a significant stage of natural language, which is "before philosophers reflect on its semantics (in particular, the semantic paradoxes)".[70] But Kripke's reason that the ascent to a metalanguage is needed, is not because the predicates 'true of exactly those predicates false or neither true nor false of themselves' are inexpressible in his language as Simmons pointed out,[71] but because semantical notion such as 'grounded', 'paradoxical', etc. cannot be expressed in his language. I would agree with Kripke's assertion that his models are plausible as models of natural language at a stage before we reflect on the generation process associated with the concept of truth, the stage which continues in the daily life of nonphilosophical speakers, on pain of giving up the goal of universal language,[72] since I think the notion of 'groundedness' and 'paradoxicality' is not in the conceptual repertoire of the ordinary speaker.

So far I explained Kripke's approach to construct a language in which truth predicate is included, and examined whether his idea is undermined by the diagonal argument. If my argument on further restriction of constructing a countervalue is right, Kripke's view is safe from the diagonal argument.

I think Kripke's advantages are these: first of all, his theory has more explanatory power than Tarski's hierarchy solution, in the respect that the language has its own truth predicate which is not different at each level: there is only one 'true' predicate. Still, in the language, Liar's paradox can be overcome by accepting truth value gap. Secondly, Kripke's solution gets

---

70) S. Kripke, ibid, p. 80.
71) K. Simmons, ibid, p. 25.
72) S. Kripke, ibid, footnote *34 p. 80.

rid of some problems of Tarski's approach such as explaining the risky character of our language and the problem of transfinite level. But there are also some defects in his theory. First his solution to the Liar is only a formal one. He did not give any account of his theory in the philosophical aspect. Second, Kripke cannot explain how our ordinary language can be a universal one without using a metalanguage. I think these defects need to be remedied to make Kripke's theory a complete one.

**Reference**

Enderton, H. Elements of Set Theory, 1977.

Haack, S. Philosophy of Logics, 1978.

Kleene, S. Introduction to Metamathmetics, 1952.

Kripke, S. "Outline of a Theory of Truth", The Journal of Philosophy, 1975. Reprinted in Recent Essays on Truth and the Liar Paradox ed. by R. L. Martin, 1984.

Simmons, K. "The Diagonal Argument and the Liar", Journal of Philosophical Logic, 1990.

Tarski, A. "The Semantic Conception of Truth", Philosophy and Phenomenological Research 4, 1944.