South American Journal of Logic
Vol. 1, n. 1, pp. 163–178, 2015
ISSN: 2446-6719

Is Tarski's Definition of Truth Circular from Frege's Point of View?

Dirk Greimann

Abstract

In Frege's system, truth is considered to be a fundamental notion of logic that cannot be defined in a non-circular way. From a modern point of view, this thesis has become obsolete. The reason is that Tarski's explicit definition of truth in terms of satisfaction is widely regarded as a successful, non-circular definition. The aim of this paper is to show that Tarski's definition is indeed circular, from Frege's point of view.

Introduction

In his first Logical Investigation, 'Der Gedanke' (1918), Frege argues that any attempt to define truth fails because truth cannot be explained in a non-circular way. He first aims to show that the classical definition of truth in terms of correspondence is circular. To this end, he distinguishes between two kinds of correspondence: correspondence in all respects ('perfect correspondence'), and correspondence in a certain specific respect. Frege rejects the definition of truth in terms of perfect correspondence because it implies that truth is a kind of identity. The definition in terms of specific correspondence, on the other hand, is criticized by him on the grounds that it is circular:

> But could we not maintain that there is truth when there is correspondence in a certain respect? But which respect? For in that case what ought we do so as to decide whether something is true? We should have to inquire whether it is true that an idea and a reality, say, correspond in the specified respect. So the attempted explanation of truth as correspondence breaks down. (Frege 1918, p. 327)

Frege then goes on to generalize this result as follows:

> And any other attempt to define truth also breaks down. For in a definition certain characteristics would have to be specified. And in application to

> any particular case the question would always arise whether it were *true* that the characteristics were present. So we should be going around in a circle. So it seems that the content of the word "true" is *sui generis* and indefinable. (Frege 1918, p. 327)

Strictly speaking, Frege's indefinability thesis is not a single thesis, but a bundle of different theses that includes, for instance, the claim that truth is *sui generis*. I shall focus here on what I take to be the core of the indefinability thesis, namely, the claim that truth is indefinable in the sense that it cannot be explained in a non-circular way: when we want to explain what truth is, we must always presuppose that this is already known.

In what follows, my aim is to show that Tarski's definition of truth in terms of satisfaction is circular in Frege's sense. In section 1, I shall briefly recapitulate Tarski's conception of truth and satisfaction. In section 2, Frege's conception of these notions is reconstructed. Finally, in section 3, it is shown that Tarski's definition of truth is circular in Frege's sense. As far as I know, the thesis defended here is not to be found in the literature.[1]

1 Tarski's conception of truth and satisfaction

The main goal of Tarski's theory of truth is to give a materially adequate and formally correct definition of the classical, Aristotelian concept of truth, which Tarski characterizes by means of the following truth-schema:

x is true if and only if p,

where 'p' is a placeholder for a sentence and 'x' a name of that sentence. A typical instance is the following sentence:

'Snow is white' is true if, and only if, snow is white.

Truth is here considered to be a semantic notion, that is, a notion describing a relation between language and the world.[2]

[1] The most important papers on this topic are Sluga 1999 and 2001, Weiner 2008, and the papers collected in Beaney/Reck 2005. The interpretation of Frege's indefinability thesis defended here is based on the reconstruction of Frege's logical system in Greimann 2000, 2008 and 2014.

[2] I shall ignore here the deflationist reading of the truth-schema that is based on the interpretation of the truth-predicate as a device of disquotation in Quine's sense. I shall also ignore Tarski's theorem of the indefinability of truth, which says that it is impossible to construct in the metalanguage an open sentence that applies to all and only the true sentences of the object language if the latter language is

For technical reasons, the truth-schema itself cannot be regarded as a formally correct truth-definition. The problem is that we cannot treat the schematic letters occurring in it as bound variables. Nevertheless, its instances can be considered as partial definitions of truth. Thus, we may consider the instance

'Snow is white' is true if, and only if, snow is white

as a satisfactory definition of the truth of this peculiar sentence (cf. Tarski 1944, p. 344).

To generalize the partial definitions, the method suggesting itself is to define truth recursively, beginning with the definition of truth for elementary sentences and then going on to define the truth for complex sentences in the well-known compositional way. Thus, we may define the truth of elementary sentences such as 'Snow is white' and 'Grass is green' by the corresponding instances of the truth-schema, and then go on to define the truth of complex sentences like 'Snow is white or grass is green' in terms of the truth of its constituent sentences.

This method is too simple, however. It does not work for quantified sentences like 'For all x: x is white'. We cannot define the truth of such sentences in terms of the truth of the sentences of which they are composed, because the constituent sentences are open sentences that do not have a truth-value at all. The constituent sentence 'x is white', for instance, is neither true nor false, because its truth-value depends on the object we choose as the value of the variable 'x'. Consequently, the notion of truth can only be applied to closed sentences, and not also to open ones.

Tarski's strategy to overcome this problem is to introduce a *more general notion* that can be applied both to open and closed sentences (cf. 1933, p. 189). This notion, which Tarski calls 'satisfaction', is characterized by him by means of a corresponding satisfaction-scheme of which the following instance is a paradigm (cf. Tarski 1933, p. 190):

The object a satisfies the open sentence 'x is white' if, and only if, a is white.

Obviously, the sentence 'Snow is white' is true if and only if snow satisfies the open sentence 'x is white'. We can hence define the truth of this sentence in terms of the satisfaction of 'x is white' by the object denoted by 'snow'. From a more traditional point of view, this amounts to defining truth in terms of the exemplification relation: the sentence 'Snow is white' is true if and only if snow exemplifies the property of whiteness, i.e., in Fregean terms, if snow falls under the concept *white*.

In order to elaborate his approach, Tarski needs to overcome two technical difficulties. The first is that there are open sentences with more than one free variable,

semantically closed (cf. Tarski 1933, p. 247). This theorem is not relevant for the reconstruction and evaluation of Frege's circularity argument.

like ‘x is the father of y’ and ‘x is more similar to y than to z’. For each number of free variables we need a corresponding relation of satisfaction. Whereas, for instance, the satisfaction of the open sentence ‘x is white’ by a given object x is a dyadic relation between an open sentence and an object, the relation of satisfaction of the open sentence ‘x is the father of y’ by two objects z and y is a triadic relation between an open sentence and two objects. The latter relation corresponds, in Frege’s logic, to the relation of the ‘standing’ of two objects in a given relation. In the case of sentences with three free variables, we need a four-adic relation of satisfaction, and so on. To define satisfaction and hence truth, we consequently need to introduce an infinite number of different relations of satisfaction.

Tarski solves this problem by means of a technical trick; he construes satisfaction as a dyadic relation between open sentences with any number of free variables and infinite sequences of objects (cf. 1933, p. 191). These infinite sequences always contain objects that are irrelevant for the satisfaction of open sentences. Nevertheless, the construction is materially adequate because it satisfies the satisfaction schema

> The sequence f satisfies the (open or closed) sentence x if and only if p,

where ‘p’ is a sentence in which the free variables occurring in the sentence denoted by ‘x’ are substituted by variables that refer to the corresponding members of the sequence (cf. 1933, p. 192).

The second and more important problem, which is not explicitly discussed by Tarski, is that the notion of satisfaction is not really a more general notion than the notion of truth in the sense that it can be applied to both open and closed sentences. We saw that the notion of truth is applicable only to closed sentences. It seems that the notion of satisfaction can only be applied to open sentences, and not also to closed ones. What are the conditions for a given object to satisfy the closed sentence ‘Snow is white’? Does grass satisfy this sentence? Or only snow? Intuitively, it does not even make sense to apply the notion of satisfaction to closed sentences.

This problem is solved by Tarski as follows. According to his satisfaction-scheme, the condition that an object must meet in order to satisfy the closed sentence ‘Snow is white’ is simply to be such that snow is white (cf. Tarski 1933, p. 192). The corresponding partial definition of satisfaction reads:

> For all objects x: x satisfies the closed sentence ‘Snow is white’ if and only if snow is white.

Since grass has the property of being such that snow is white, it satisfies the closed sentence ‘Snow is white’ as well. If one object satisfies this condition, then any other object satisfies this condition, too.[3] Hence, either all objects satisfy a given closed sentence or none. This lemma, which is Tarski's Lemma B, allows us to define truth as a special case of satisfaction, namely, as the satisfaction of a closed formula by all objects (cf. Tarski 1933, pp. 195, 198).

It is obvious that Tarski's atomistic approach to define truth in terms of satisfaction is highly artificial. It would be more natural to follow the converse strategy of defining satisfaction and kindred relations like exemplification and the falling of an object under a concept in terms of truth. Thus, we could give the following partial definition of satisfaction in terms of truth, as also Tarski observes (cf. Tarski 1944, p. 353):

> The object snow satisfies the open sentence 'x is white' if, and only if, the closed sentence 'Snow is white' is true.

We shall see that Frege's conception of truth is characterized by this approach.

2 Frege's Approach

In Tarski's system, satisfaction is construed as a relation between objects and sentences (or 'sentential functions'). The term 'satisfaction' is used here in a metaphorical sense. In the literal sense, an object may be said to satisfy a *condition*, but not a sentence. Thus, snow satisfies the condition of being white. When we say that snow satisfies the sentence 'x is white', what we mean is that snow satisfies the condition expressed by this sentence. The Tarskian notion of satisfaction is hence a derivative notion.

Frege does not use the term 'satisfaction' at all, of course. Instead, he speaks of the falling of an object under a concept and the standing of two objects in a relation (see, for instance, Frege 1893, §4). In a broad sense, such relations can also be regarded as relations of satisfaction. Thus, snow satisfies the concept *white* in the derivative sense that it satisfies the condition of falling under this concept. The same applies to the relation of the exemplification of properties and relations by objects.

Let us call a conception of truth *atomistic* when it explains truth in terms of exemplification or kindred relations like Tarski's relation of satisfaction. A conception of truth is *holistic* when it explains, conversely, such relations in terms of truth. Since Tarski explains truth in terms of satisfaction, his conception of truth is atomistic. Frege's conception, on the other hand, is holistic: he explains the logical relation of an object falling under a concept and kindred relations in terms of truth, and not *vice versa*. According to Frege's first system, which is formulated in the *Begriffsschrift* (1879),

[3] This condition is identical to the condition that an object must satisfy in order to be an element of the set $\{x \mid \text{ snow is white}\}$. Since it is true that snow is white, any object satisfies this condition. If it is true that p, then $\{x \mid \text{ p}\}$ is identical to the universal set, and, if it is false that p, then $\{x \mid \text{ p}\}$ is identical to the empty set.

snow falls under the concept white if and only if the whiteness of snow is a fact, and, according to the second system, which is described in *Grundgesetze* (1893), snow falls under the concept *white* if and only if the truth-value of snow's being white is the True. In what follows, this is explained in more detail.

2.1 Truth and satisfaction in Frege's first system

In the first paragraphs of *Begriffsschrift*, Frege rejects the traditional distinction between the subject and the predicate of a judgement on the ground that sentences like 'Snow is white', 'Whiteness is a property of snow' and 'The whiteness of snow is a fact' have exactly the same logical content though they have different grammatical subjects. To overcome this problem, he replaces the distinction between subject and predicate with the more flexible distinction between function and argument (cf. §3 and §9). In 'Whiteness is a property of snow', both snow and whiteness can be regarded as the argument. In the first case, 'Whiteness is a property of x' is the corresponding function and, in the second case, 'Φ is a property of snow'.

In a logically transparent language, the sentences must be formulated in such a way that the distinction between subject and predicate cannot be applied. To achieve this, Frege translates the predicates of natural language into his formal language as functional signs. In §3 of *Begriffsschrift*, he describes the syntax of his formal language (called 'the Begriffsschrift') as follows:

> Imagine a language in which the sentence 'Archimedes was killed at the capture of Syracuse' is expressed in the following way: 'The violent death of Archimedes at the capture of Syracuse is a fact'. Even here, if one wants, subject and predicate can be distinguished, but the subject contains the whole content, and the predicate serves only to present it as a judgement. *Such a language would have only one predicate for all judgements, namely, 'is a fact'.* [...] *Our Begriffsschrift is such a language and the symbol* —— *is its common predicate for all judgements.* (Frege 1879, §3, partly my translation)

According to this description, the formal language contains a single syntactic predicate, namely, 'x is a fact'. A sentence like 'Snow is white' is translated into the formal language as 'The whiteness of snow is a fact', where 'the whiteness of x' is the functional sign corresponding to the predicate 'x is white'. The sentence 'Romeo loves Juliet' must accordingly be translated as 'The love of Romeo to Juliet is a fact', and so on.

In accordance with the treatment of predicates as functional signs, Frege treats the properties and relations denoted by predicates as functions. Thus, the property of being white is construed by him as the function *the whiteness of x*, which is an 'unsaturated' or 'incomplete' propositional content lacking an object that fills out the empty place marked by 'x'. This incomplete content, the *whiteness of x*, is a function mapping snow onto the whiteness of snow, that is, the propositional content that snow is white.

As Frege points out, 'is a fact' is not to be regarded as a *genuine* predicate, that is, it does not express the predicative part of a propositional content. In 'The whiteness of snow is a fact', the subject 'The whiteness of snow' already contains the whole propositional content, and 'is a fact' serves only 'to present this content as a judgement', i.e., to assert this content is true. The same applies, *mutatis mutandis*, to the assertion sign '——'. This symbol is, in the first place, an illocutionary act indicating device, and, in the second place, a truth-operator.

In §2, Frege explains the difference between '—— A' and '— A' in more detail: whereas '—— A' expresses a judgement, '— A' expresses the content A without expressing a judgement. This distinction is designed to make explicit that we may express a propositional content without asserting it as true. The syntactic structure of the assertoric sentences of natural language is logically misleading, in Frege's view, because it hides the fact that an assertion consists of two logically independent acts: the expression of a propositional content, and the assertion of its truth (cf. Frege 1897, p. 239).

To represent the duality of expressing a thought and asserting its truth syntactically in his formal language, Frege needs to construct a category of expressions that express a propositional content without asserting it as true. He achieves this, again, by replacing the predicates by functional signs. Thus, the sentence-nominalization 'the whiteness of snow' expresses that snow is white without asserting this content as true. To assert something as true, we need to apply the judgement-stroke: '—— White(snow)' does not only express a propositional content, but also its truth.

Note that, strictly speaking, the judgment-stroke does not correspond to the predicate 'is a fact', but to the 'form of the assertoric sentence', that is, the syntactic type that distinguishes assertoric sentences from interrogative and imperative ones. This form is an illocutionary act indicating device that expresses, in the first place, the assertoric force with which we normally utter assertoric sentences. In the second place, it expresses also the truth of propositional contents. Thus, the sentence 'Snow is white' expresses, in virtue of its assertoric form, that its content is true. The interrogative sentence 'Is snow white?' expresses exactly the same content, but without expressing its truth. The same applies to sentence-nominalizations like 'the whiteness of snow'.

In his later writings, Frege says explicitly that it is by means of the assertoric form that we express that a given content is true. In the manuscript entitled 'Logik' (1897), he writes that

> ...it is really by using the form of the assertoric sentence that we express truth [*womit wir die Wahrheit aussagen*], and to do this we do not need the word 'true'. Indeed, we can say that even where we use the locution

> ‘it is true that ...’ the essential thing is really the form of the assertoric sentence. (Frege 1897, p. 229, partly my translation)

He reaffirms this thesis in a passage from a posthumous writing entitled ‘Logic in Mathematics’ from 1914:

> In order to put something forward as true, we do not need a special predicate: we need only the assertoric force with which the sentence is uttered. (Frege 1914, p. 233)

The insight behind this view that, by asserting a content, we are presenting it as a fact. For Frege, to judge is always to judge something as true, and to assert is always to assert something as true. The judgement-stroke represents the truth-claim that is implicitly contained in judgements and assertions.[4]

Note that the functional expression ‘— White(snow)’ (or ‘the whiteness of snow’) does not express the falling of snow under the concept *white*, but only the result of the saturation of this concept by this object. It leaves open whether this result is a fact or not. The brackets express only the saturation (or ‘completion’) of the function by the argument, but not the falling of the argument under the corresponding concept. To express that snow falls under the concept white, we must also express that White(snow) is a fact. To this end, we need the judgement-stroke: ‘—— White(snow)’ expresses not only that snow is white, but also that this is a fact.

Now, in §10 of *Begriffsschrift*, Frege introduces the logical relations of satisfaction of his system:

> In order to express an indeterminate function of the argument A, let us enclose A in brackets after a letter, as in
>
> $$\phi(\mathrm{A}).$$
>
> Similarly,
>
> $$\Psi(\mathrm{A,B})$$
>
> signifies a function of the two arguments A and B. [...]
>
> $$\text{——}\ \Phi(\mathrm{A})$$
>
> can be read: ‘A has the property Φ’.
>
> $$\text{——}\ \Psi(\mathrm{A,B})$$
>
> may be translated as ‘B stands in the Ψ-relation R to A’ or ‘B is the result of an application of the procedure Ψ to A’.

[4] For a detailed analysis of Frege’s notion of assertion and assertoric force, see Greimann 2000 and especially 2012.

Given our reconstruction of the judgement-stroke as an illocutionary truth-operator, Frege considers here the satisfaction of concepts and relations by objects as special cases of truth. To say that an object A has the property Φ means to say that A's being Φ is a fact. In Frege's second system, this holistic approach becomes even more evident, as we shall see.

2.2 Truth and satisfaction in Frege's second system

Frege's second logical system is characterized by the distinction of sense and reference. We saw that, in his first system, an ordinary sentence like 'Snow is white' is paraphrased as 'The whiteness of snow is a fact'. From the point of view of his second system, this translation involves a confusion of sense and reference. While the reference of 'Snow is white' is a truth-value, the reference of the nominalization 'The whiteness of snow is a fact' is the sense expressed by 'Snow is white'.

To separate the act of expressing a thought from the act of asserting it as true, we must hence construct sentence-nominalizations that express the senses of sentences and denote their truth-values. Since, in his second system, Frege construes sentences as names of truth-values, the sense of a sentence can be considered as a *mode of presentation* of a truth-value. Thus, the sense of 'Snow is white' is identical to the sense of the proper name 'the truth value of: that snow is white'. Note that the sentence 'Snow is white' and the proper name 'the truth value of: that snow is white' have exactly the same truth-conditions: both denote the True if and only if snow is white. For this reason, they also express the same sense (cf. Frege 1893, §32). The difference is that the proper name expresses this sense without asserting it as true.

In his second system, Frege uses the predicates of natural language as functional signs, again. Because of the distinction between sense and reference, the functions denoted by these signs are this time not considered as functions mapping objects onto propositional contents, but as functions mapping objects onto truth values. That is, a predicate like 'x is white' is used in the formal language as a functional sign designating the function *the truth-value of: that x is white.* This function maps a given object x onto the True if and only if x is white.

As a consequence of this use, the predicates of natural language cannot be used in the formal system to assert anything. Considered as an expression of the formal language, '$2 + 3 = 5$' is not an assertoric sentence, but a proper name of a truth-value whose counterpart in natural language is the definite truth-value description 'the truth-value of: that 2+3 is identical to 5'. Such a nominalization expresses a thought and designates a truth-value, but it does not assert the thought as true. To assert something, we hence need a special sign that allows us to express that a given truth-value is the True. This is the task of the judgement-stroke. Whereas '$2+3=5$' simply designates a certain truth-value, without saying which of the two it is, the assertoric sentence '—— $2 + 3 = 5$' expresses that the truth-value designated by '$2 + 3 = 5$' is the True. These features of the formal language are clearly described by Frege in the following passage from *Funktion und Begriff* (1891):

> If we write down an equation or inequality, e.g. $5 > 4$, we ordinarily wish at the same time to express a judgement; in our example, we want to assert that 5 is greater than 4. According to the view I am here presenting [i.e. the treatment of predicates as functional signs, D.G.], '$5 > 4$' and '$1 + 3 = 5$' just give us expressions for truth-values, without making any assertion. [...] We thus need a special sign in order to be able to assert something. To this end I make use of a vertical stroke at the left end of the horizontal, so that, e.g., by writing '—— $2 + 3 = 5$' we assert that 2+3 equals 5. Thus here we are not just writing down a truth-value, as in '$2 + 3 = 5$', but also at the same time saying that it is the True. (1891, p. 142).

Frege says here almost explicitly that, in '—— $2+3 = 5$', the judgement-stroke expresses that the truth-value of: that $2 + 3 = 5$ is *the True*. Hence, the judgement-stroke works as a kind of illocutionary truth-operator, again.[5] It expresses that a given truth-value is the True.

Now, in §4 of *Grundgesetze*, the satisfaction of concepts and relations by objects is explained as follows:

> We say that the object Γ stands in the relation $\Psi(\xi, \zeta)$ to the object Δ if $\Psi(\Gamma, \Delta)$ is the True, just as we say that the object Δ falls under the concept $\Phi(\xi)$ if $\Phi(\Delta)$ is the True.

Clearly, Frege explains here satisfaction in terms of truth, and not *vice versa*. This holistic order of explanation is a consequence of the treatment of concepts and relations as functions whose value is always a truth-value. It implies that the falling of an object Δ under the concept Φ consists in the identity of $\Phi(\Delta)$ with the True. The standing of the objects Γ and Δ in the relation Ψ consists, accordingly, in the identity of $\Psi(\Gamma, \Delta)$ with the True.

Note that, in order to express that snow falls under the concept *white*, it does not suffice to apply the functional sign 'White(x)' to 'snow'; it is, in addition, necessary to express that the result of this application designates the True. The reason is that the functional expression 'White(snow)' designates the truth-value of: that snow is white, without saying which of the two it is. To express that snow actually falls under the concept *white*, we must express that the truth-value of: that snow is white, is the True. To this end, we must apply the judgement-stroke to 'White(snow)'.

---

[5] For a detailed justification of this reconstruction, see Greimann 2000, 2008, and 2014.

In §5 of *Grundgesetze*, Frege shows that the judgement-stroke cannot be reduced to any other expression of the formal language. Assume that we want to assert '$2+3=5$' as true. Although '$(2=2)$' denotes the True, Frege argues, the expression '$(2+3=5)=(2=2)$' cannot be used to assert '$2+3=5$' as true. The reason is that the complex name '$(2+3=5)=(2=2)$' merely designates a truth-value, without saying which of the two it is. This truth-value is the truth-value of: that (the truth-value of: that $2+3=5$) is identical to (the truth-value of: that $2=2$). The problem is that '$(2+3=5)=(2=2)$' does not say whether the truth-value designated by it is the True or the False. For this reason, '$(2+3=5)=(2=2)$' leaves open what the truth-value of '$2+3=5$' is supposed to be. As a consequence, '$(2+3=5)=(2=2)$' cannot be used to assert '$2+3=5$' as true, even when we presuppose that the hearer knows that '$2=2$' denotes the True. Frege writes:

> We have already said above that nothing at all is asserted in a mere equation; '$2+3=5$' simply designates a truth-value, without saying which of the two it is. Even if I wrote '$(2+3=5)=(2=2)$' and presupposed that it was known that $2=2$ is the True, I have not thereby asserted that the sum of 2 and 3 is 5, but have merely designated the truth-value of: '$2+3=5$' refers to [*bedeute*] the same thing as '$2=2$'. We therefore need another special sign to be able to assert something as true. For this purpose I place before the name of the truth-value the sign '——' [...]. (Frege 1893, p. 215)

This implies that the judgement-stroke is a sign *sui generis* in the sense that it has a very special expressive function in virtue of which it differs from any other expression of the formal language.

3 The Circularity of Tarski's Definition of Truth

Frege's holistic approach to explain satisfaction in terms of truth implies that truth cannot be explained in a non-circular way. The circle in question is more clearly described in the posthumous writing 'Logik' from 1897:

> Now it would be futile to employ a definition in order to make it clearer what is to be understood by 'true'. If, for example, we wished to say 'an idea is true if it agrees with reality' nothing would have been achieved, since, in order to apply this definition we should have to decide whether some idea or other did agree with reality. Thus we should have to presuppose the very thing that is being defined. The same would hold of any definition of the form 'A is true if and only if it has such-and-such properties or stands in such-and-such a relation to such-and-such a thing'. In each case in hand it

> would always come back to the question whether it is true that A has such-and-such properties, or stands in such-and-such a relation to such-and-such a thing. Truth is obviously something so primitive (*etwas so Ursprüngliches*) and simple (*Einfaches*) that it is not possible to reduce it to anything still simpler. (Frege 1897, p. 228)

Frege's argument seems to be this. Let D be a definition in which truth is defined in terms of the property (or 'characteristic') P. The definition may read:

x is true if, and only if, x is P.

Assume that we want to apply D to decide whether a given sentence S is true. To this end, we must decide whether S is P. This leads to a circle because the question arises whether *it is true* that S is P. To decide whether S is true, we must decide whether it is true that S is P.

This circle is a consequence of Frege's analysis of satisfaction in terms of truth.[6] To decide whether a given sentence S is P, we must decide whether S falls under the concept P. Since concepts are functions whose value is always a truth-value, we must decide, to this end, whether the truth-value of: that S is P, is the True. It is in this sense that, in order to decide whether S is P, we must decide whether *it is true* that S is P. To subsume S under P, we need to identify the truth-value of the functional application P(S) with the True. This, finally, implies that any definition of truth is circular in the sense that its application to a special case already presupposes that we know what truth (or 'the True') is.

Obviously, this argument also applies to Tarski's definition of truth, according to which a sentence is true if and only if it is satisfied by all infinite sequences of objects. To decide, in a particular case, whether a given sentence has this property, we must identify the value of the corresponding function for this sentence as argument with the True.

It is, however, important to note that Frege's indefinability thesis does not refer to the Tarskian truth-predicate 'x is a true sentence', but to the illocutionary truth-operator, i.e., the form of the assertoric sentence or the judgement-stroke, respectively. Whereas the *definiendum* of Tarski's definition of truth is the predicate 'x is a true sentence', the *definiendum* of the truth-definitions discussed by Frege is the form of the assertoric sentence, considered as the effective truth-operator of natural language. Consequently, Frege's thesis does not imply that Tarski's semantic notion of truth is

[6] In the literature, it is often assumed that this is a consequence of Frege's thesis that sentence-pairs of the form 'p' and 'It is true that p' express the same sense. This reconstruction, however, implies that Frege's argument is a *non-sequitur*, because the circle described by him is not vicious. For more details on this, see Soames 1999, Chap. 2.

indefinable, but only that this notion cannot be used to define the notion of truth that is associated with the form of the assertoric sentence.

For Frege, the notion of truth that is basic for logic is not the semantic notion, but the notion corresponding to the form of the assertoric sentence. This becomes clear from the following passage from the posthumous writing 'Meine grundlegenden logischen Einsichten' (1915):

> So the word 'true' seems to make the impossible possible: namely, to allow what corresponds to the assertoric force to assume the form of a contribution to the thought. And although this attempt miscarries, or rather through the very fact that it miscarries, it indicates what is characteristic of logic. And this, from what we have said, seems something essentially different from what is characteristic of aesthetics and ethics. For there is no doubt that the word 'beautiful' actually does indicate the essence of aesthetics, as does 'good' that of ethics, whereas 'true' only makes an abortive attempt to indicate the essence of logic, since what logic is really concerned with is not contained in the word 'true' at all but in the assertoric force with which a sentence is uttered. (Frege 1915, p. 323)[7]

Frege suggests here that logic is concerned with the notion of truth that is associated with the speech act of assertion. His indefininability thesis refers, accordingly, to the judgement-stroke. For Tarski, on the other hand, the notion of truth that is basic for logic is the semantic one. His definition of truth refers, accordingly, to the semantic notion. Hence, Frege's argument does not show that Tarski's definition of truth is circular; it only shows that this definition cannot be used to define Frege's notion of truth in a non-circular way.

In Frege's system, Tarski's semantic notion of truth is expressed by the predicate 'x denotes (*bedeutet*) the True', which does not belong to the formal language, but to the meta-language (cf. Frege 1893, §§2 ff.). A sentence denotes the True if and only if it is true in Tarski's sense. According to §10 of *Grundgesetze*, the Horizontal '— x', which may be read as 'x is the True', can be reduced to the identity sign. Roughly speaking, the corresponding definition reads:

— x if, and only if, x = (x = x).

That is, x is the True if and only if x is identical to the truth-value of: that x = x. The predicate 'x is the True' and hence also the predicate 'x denotes the True' are thus definable in Frege's system. This confirms, again, that his indefinability thesis

[7]For a reconstruction of Frege's puzzling remarks in this passage, see Tascheck 2008 and also the alternative reconstruction in Greimann 2012 and 2014.

does not refer to the semantic notion of truth, but to the notion associated with the judgement-stroke.[8]

References

Note: The references to Tarski's and Frege's writings refer to the English translations.

Beaney, Michael, and Reck, Erich (eds.) (2005). *Gottlob Frege. Critical Assessment of Leading Philosophers*, Vol. II, *Frege's Philosophy of Logic*. London and New York: Routledge.

Frege, Gottlob (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, reprinted in his *Begriffsschrift und andere Aufsätze*, Hildesheim: Olms, second edition, 1988, pp. V–88. A selection is translated as 'Begriffsschrift' in Frege 1997, pp. 47–78.

Frege, Gottlob (1891). *Funktion und Begriff*. Jena: Hermann Pohle, reprinted in: Frege 1990, pp. 125–42. Translated as 'Function and Concept' in Frege 1997, pp. 130–148.

Frege, Gottlob (1893). *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, Vol. 1. Jena: Hermann Pohle. Reprint: Darmstadt: Wissenschaftliche Buchgesellschaft, 1962. A selection is translated as 'Grundgesetze de Arithmetik, Volume I, Selections' in Frege 1997, pp. 194–233.

Frege, Gottlob (1897). 'Logik'. Posthumous writing, in: Frege 1983, pp. 137–163. An extract is translated as 'Logic: Extract' in Frege 1997, pp. 227–250.

Frege, Gottlob (1914). 'Logik in der Mathematik' Posthumous writing, in: Frege 1983, pp. 219–70. Translated as 'Logic in Mathematics' in Frege 1979, pp. 203–250.

Frege, Gottlob (1915). 'Meine grundlegenden logischen Einsichten'. Posthumous writing, in: Frege 1983, pp. 271–272. Translated as 'My Basic Logical Insights' in Frege 1997, pp. 322–324.

Frege, Gottlob (1918). 'Der Gedanke. Eine logische Untersuchung'. *Beiträge zur Philosophie des deutschen Idealismus*, 1, pp. 58–77. Reprinted in: Frege 1990, pp. 342–362. Translated as 'Thought' in Frege 1997, pp. 325–345.

[8] I am grateful to Rodrigo Cezar Medeiros Moreira and Heitor Achilles for their helpful comments on an earlier draft of this paper.

Frege, Gottlob (1979). *Posthumous Writings*. Ed. by H. Hermes, F. Kambartel and F. Kaulbach, and translated by P. Long and R. White, Oxford: Basil Blackwell. Engl. trans. of the first edition of Frege, *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, Band 1, ed. by H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969.

Frege, Gottlob (1983). *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, Vol. 1. Ed. by H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, second and extended edition (first published in 1969). Translated as *Posthumous Writings* in Frege 1979.

Frege, Gottlob (1990). *Kleine Schriften*. Ed. by I. Angelelli, second edition, Hildesheim: Olms.

Frege, Gottlob (1997). *The Frege Reader*. Ed. and introduced by Michael Beaney, Oxford: Blackwell.

Greimann, Dirk (2000). 'The Judgement-Stroke as a Truth-Operator: A New Interpretation of the Logical Form of Sentences in Frege's Scientific Language', *Erkenntnis* 52, pp. 213–238.

Greimann, Dirk (2008). 'Does Frege Use a Truth-Predicate in His Justification of the Logical Laws?', *Mind* 117, pp. 403–425.

Greimann, Dirk (2012). 'A Caracterização da Lógica pela Força Assertórica em Frege. Resposta a Marco Ruffino', *Manuscrito* 35, pp. 61–83.

Greimann, Dirk (2014). 'Frege on Truth, Assertoric Force and the Essence of Logic', *History and Philosophy of Logic* 35, pp. 1–17.

Sluga, Hans (1999). 'Truth before Tarski'. In: *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empiricism*, ed. by J. Wolénski and E. Köhler, Dordrecht, Kluwer, pp. 27–41.

Sluga, Hans (2001). 'Frege on the Indefinability of Truth'. In: *From Frege to Wittgenstein*, ed. by E. Reck, Oxford, Oxford University Press, pp. 75–95.

Soames, Scott (1999). *Understanding Truth*, New York, Oxford University Press.

Tarski, Alfred (1933). 'O pojeciu prawdy w odniesieniu do sformalizowanych nauk dedukcyjnych'. *Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno-Fizycznych* 34, Warsaw. Translated as 'The Concept of Truth in Formalized Languages'. In: A. Tarski, *Logic, Semantics, Metamathematics:* Papers from 1923 to 1938. Oxford: Clarendon Press, 1956, pp. 152–278.

Tarski, Alfred (1944). 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* 4, pp. 341–375.

Taschek, William (2008). 'Truth, Assertion, and the Horizontal: Frege on 'The Essence of Logic'', *Mind*, 117, pp. 375–401.

Weiner, Joan (2008). 'How Tarskian is Frege?'. *Mind* 117, pp. 427–450.

Dirk Greimann
CNPq and Department of Philosophy
Universidade Federal Fluminense (UFF)
Rua Professor Marcos Waldemar de Freitas Reis, s/n, CEP 24210-201, Niteroi, RJ, Brazil
E-mail: greimann@dirk-greimann.de