Journal for the History of Analytical Philosophy

Volume 3, Number 10

Editor in Chief
Kevin C. Klement, University of Massachusetts

Editorial Board
Gary Ebbs, Indiana University Bloomington
Greg Frost-Arnold, Hobart and William Smith Colleges
Henry Jackman, York University
Sandra Lapointe, McMaster University
Lydia Patton, Virginia Tech
Marcus Rossberg, University of Connecticut
Mark Textor, King's College London
Audrey Yap, University of Victoria
Richard Zach, University of Calgary

Review Editors
Juliet Floyd, Boston University
Chris Pincock, Ohio State University

Assistant Review Editor
Sean Morris, Metropolitan State University of Denver

Design
Daniel Harris, Hunter College

© 2015 Monika Gruber

Carnap's Contribution to Tarski's Truth

Monika Gruber

In his seminal work "The Concept of Truth in Formalized Languages" (1933), Alfred Tarski showed how to construct a formally correct and materially adequate definition of *true sentence* for certain formalized languages. These results have, eventually, been accepted and applauded by philosophers and logicians nearly in unison. Its *Postscript*, written two years later, however, has given rise to a considerable amount of controversy. There is an ongoing debate on what Tarski really said in the postscript. These discussions often regard Tarski as putatively changing his logical framework from type theory to set theory.

In what follows, we will compare the original results with those presented two years later. After a brief outline of Carnap's program in *The Logical Syntax of Language* we will determine its significance for Tarski's final results.

Carnap’s Contribution to Tarski’s Truth

Monika Gruber

In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments.
(“The Principle of Tolerance”, Carnap 1937, 52)

1. Introduction

Although it was common practice for mathematicians to employ semantical notions in their work, no coherent theory of these notions existed at the beginning of the twentieth century. Semantical notions had not been successfully defined and there were no axiomatic theories taking them as primitive either. In 1933, Tarski wrote his most famous article “The Concept of Truth in Formalized Languages” (CTFL) with precisely this goal: to define semantical notions in such a way that would make them mathematically acceptable in the main logical frameworks—type theory and set theory.

The most prominent result of this work (1933) is a formally correct and materially adequate definition of *true sentence*. With his famous meticulousness, Tarski showed how to define *true sentence* for formalized languages of finite order. In the last chapter, using as an example the language of the general theory of classes, he claimed it was impossible to construct such a semantical definition for the languages of infinite order. Two years later, in 1935, the German translation of the paper appeared. It included an additional chapter, *Postscript*, written by Tarski. At the beginning of the postscript Tarski admits that he can no longer agree with all of the results he had reached a couple of years ago. In particular, in addition to the previously studied languages, he decides to investigate the languages the structure of which cannot be brought into harmony with the principles of the theory of semantical categories. Following this, he presents a method of constructing a definition of *true sentence* for all formalized languages, including the languages of infinite order. Ever since the publication of the German, and even more so of the English translation, there has been a vivid discussion concerning the interpretation of the postscript.

While many philosophers applauded Tarski’s ingenious strategy of extending the scope of the application of the presented method of defining truth to all languages, others, specifically deflationists, had their reservations and claimed that Tarski did no such thing.[1] The debate on the postscript is still open and there is no consensus among Tarski’s readers.

In Section 2, we present Tarski’s original results and discuss the reasons for his failure to define truth for formalized languages of infinite order within his interpretation of the simple theory of types. In Section 3, we consider Carnap’s *The Logical Syntax of Language* (LSL) and determine whether and in what way it contributed to the positive conclusions Tarski reached in the postscript. In Section 4, we present an outline of Tarski’s new method and determine how it allowed him to arrive at the final conclusions of the postscript. The sections will be presented in a chronological order which should enable us to follow the development of Tarski’s methods from the original CTFL, through Carnap’s LSL, towards the postscript. In the conclusion, we summarize Tarski’s results and the relevance of Carnap’s method, in this way contributing to the debate on

[1] Among the philosophers who accepted Tarski’s extending of the method of defining truth to all formalized languages, perhaps the most prominent supporters are Anil Gupta and Nuel Belnap (1993), Saul Kripke (1975) and Donald Davidson (1984). Those opposing Tarski’s program are naturally the deflationists: Paul Horwich (1998), and Hartry Field, who makes an explicit argument against Tarski in *Saving Truth from Paradox* (2008, 33–6).

Tarski's postscript.

2. Tarski 1933

Wahrheitsbegriff: Original Framework and Results

Working within his interpretation of the simple theory of types (STT), the concept of *semantical category* was of crucial importance for Tarski's investigations. The concept, first used by Husserl, was introduced into formal sciences by Leśniewski. There are clear parallels between Tarski's theory of semantical categories and Russell and Whitehead's theory of types, although Tarski emphasizes that from the formal point of view his theory resembles rather Chwistek's simplified theory of types, and is even an extension of Carnap's *Typentheorie* presented in *Abriss der Logistik* (cf. Tarski 2006, 215). Tarski does not present us with a definition of the notion of semantical category, but instead he explains that

> . . . two expressions *belong to the same semantical category* if (1) there is a sentential function which contains one of these expressions, and if (2) no sentential function which contains one of these expressions ceases to be a sentential function if this expression is replaced in it by the other. (Tarski 2006, 216)

Moreover, it is intuitively clear to Tarski that "in order that two expressions shall belong to the same semantical category, it suffices if there exists *one* function which contains one of these expressions and which remains a function when this expression is replaced by the other" (Tarski 2006, 216). Tarski called it the *first principle of the theory of semantical categories* and it led him directly to formulating the law regarding the semantical categories of sentence-forming functors, i.e., signs representing sentential functions: "the functors of two primitive sentential functions belong to the same category if and only if the number of arguments in the two functions is the same, and if any two arguments which occupy corresponding places in the two functions also belong to the same category" (Tarski 2006, 217).

Depending on the multiplicity of the semantical categories appearing in the language and on whether the variables of the language belong to a finite or an infinite number of categories and in the latter case whether the orders of these categories are bounded above or not, Tarski distinguishes four kinds of languages:

1. languages in which all variables belong to one and the same semantical category (e.g. the calculus of classes, the sentential calculus + $\forall$, $\exists$)
2. languages in which the number of categories in which the variables are included is greater than 1 but finite (the variables are bounded above, e.g. the language of the logic of two-termed relations)
3. languages in which the variables belong to infinitely many different categories but the order of these variables does not exceed a previously given natural number *n* (the variables are bounded above, e.g. the language of the logic of many-termed relations)
4. languages which contain variables of arbitrarily high order (the variables are not bounded above, e.g. the language of the general theory of classes)

Languages of the first three kinds, in which the variables are bounded above, are *languages of finite order*, in contrast to languages of the fourth kind, in which the variables are *not* bounded above, which are *languages of infinite order*.

In writing the original version of the article (1933), Tarski had in mind only formalized languages the structure of which adheres to the theory of semantical categories. The notion of *order of semantical category* played a crucial role for the languages investigated in the Polish original.

> We require a classification of the semantical categories; to every category a particular natural number is assigned called the *order*

> *of the category.* This order is also assigned to all expressions which belong to this category. The meaning of this term can be determined recursively. For this purpose we adopt the following convention (in which we have in mind only those languages which we shall deal with here and we take account only of the semantical categories of the variables): (1) the 1st order is assigned only to the names of individuals and the variables representing them; (2) among expressions of the $n+1th$ order, where n is any natural number, we include the functors of all those primitive functions all of whose arguments are of at most the *nth* order, where at least one of them must be of exactly the *nth* order. Thanks to the above convention all expressions which belong to a given semantical category have the same order assigned to them, which is therefore called the order of that category. (Tarski 2006, 218)

Thus, the 1st order includes only the names of individuals and the variables representing them. To the 2nd order belong the names of classes of individuals and the names of two-, three-, and many-termed relations between individuals. The $(n+1)th$ order is assigned to the functors of all primitive functions all of whose arguments are of at most the *nth* order (at least one of them is exactly of the *nth* order). It is important to notice that the orders of the variables occurring in a language determine the order of this language. Furthermore, Tarski defines the notion of *semantical type* which depends on the number of free variables of a given semantical category, i.e. if the number of free variables of every semantical category in two functions is the same, then these functions are of the same semantical type (cf. Tarski 2006, 219).

Defining truth for the languages of the 1st kind did not present many difficulties for Tarski. By means of the concept of *satisfaction* of a sentential function by a sequence of objects, introduced in §3, he was able to define the concept of *true sentence* for the language of the calculus of classes. Thus, since we are considering sentences, i.e. sentential functions with no free variables, every infinite sequence of classes must satisfy a given sentence if it is to be true. One of the examples given by Tarski is: *every infinite sequence of classes satisfies the function* $x_1 \subseteq x_1$*, hence* $\forall x_1(x_1 \subseteq x_1)$ *is a true sentence.*

Trying to define truth for the languages of the 2nd, the 3rd, and the 4th kinds, Tarski introduced two new methods, which he called the *method of many-rowed sequences* and the *method of semantical unification*. Without going into details, we will just note that the application of the method of many-rowed sequences requires satisfaction to be seen not as a two-termed relation, as in the case of the language of the calculus of classes, but as a three-termed relation holding between sequences of individuals, sequences of two-termed relations and sentential functions (Tarski 2006, 227). Tarski used this method in order to define truth for the 2nd kind of languages. Thus, a new mode of expression is used: "the sequence f of individuals and the sequence F of relations together satisfy the sentential function x". Therefore:

> ... the sequence f of individuals and the sequence F of relations together satisfy the function $\rho_{1,2,3}$ if and only if the individual f_2 stands in the relation F_1 to the individual f_3 ... (Tarski 2006, 227)

He introduced the second method, of semantical unification, stating that it can be successfully applied to languages of both the 2nd and the 3rd kind. The method of semantical unification requires that a category unifying all the variables of the languages be introduced, which itself cannot be of lower order than any of the variables of the language. Consequently, sequences of the terms of this category and the relation of satisfaction holding between these sequences and the corresponding sentential functions must be of higher order than all the variables of the language. However, in the 4th kind of languages the variables are of arbitrary high order, which means that there is an "infinite diversity" of semantical categories in the language, which excludes the method of many-rowed sequences, and if we wanted to apply the method of semantical unification we would have to use expressions of infinite order, which were not available in the languages the structure of which adheres to the theory of semantical categories (cf. Tarski 2006, 243–44).

Therefore, it is clear that Tarski's commitment to the STT was the reason for reaching the negative results in regard to the languages of infinite order. For the sake of clarity, we remind the reader of the three final theses with which Tarski closed the original Polish version of his paper:

A. *For every formalized language of finite order a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage, making use only of expressions of a general logical kind, expressions of the language itself as well as terms belonging to the morphology of language, i.e. names of linguistic expressions and of the structural relations existing between them.*

B. *For formalized languages of infinite order the construction of such a definition is impossible.*

C. *On the other hand, even with respect to formalized languages of infinite order, the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concepts of the metalanguage and determining its fundamental properties by means of the axiomatic method* (the question whether the theory of truth established in this way contains no contradiction remains for the present undecided). (Tarski 2006, 265–66)

3. Carnap 1934

Logical Syntax of Language

Simultaneously to Tarski's investigations of a semantical definition of truth in formalized languages, Rudolf Carnap developed a syntactical method of defining parallel concepts. Carnap's *The Logical Syntax of Language* is undoubtedly one of the most outstanding contributions to the development of analytic philosophy in the twentieth century. Its significance has not unjustly been compared with that of Tarski's CTFL. Perhaps due to the fact that Carnap's ingenious ideas have not always been granted the appreciation they deserve, neither by his contemporaries nor by the future generations, the parallels between LSL and CTFL have often been overlooked. Nevertheless, the excellence of Carnap's method is indisputable as are its parallels to Tarski's concept of truth. To better understand the significance of Carnap's program for Tarski's truth definition, let us take a look at the essential similarities and differences between CTFL and LSL. Carnap sees the goal of LSL as

> . . . an attempt to provide, in the form of an exact syntactical method, the necessary tools for working out the problems of the logic of science. This is done in the first place by the formulation of the syntax of two particularly important types of language which we shall call, respectively, 'Language I' and 'Language II' (Carnap 1937, xiii)

Carnap distinguishes between *word-languages*, today usually referred to as *colloquial languages*, and *symbolic languages*, i.e. formal languages. Carnap regards both Languages I and II as *calculi*, hence they are what he calls *symbolic languages*. In the first introductory sentence Carnap defines the *logical syntax* of a language as

> . . . the formal theory of the linguistic forms of that language—the systematic statement of the formal rules which govern it together with the development of the consequences which follow from these rules.
>
> A theory, a rule, a definition, or the like is to be called *formal* when no reference is made in it either to the meaning of the symbols (for example, the words) or to the sense of the expressions (e.g. the sentences), but simply and solely to the kinds and order of the symbols from which the expressions are constructed. (Carnap 1937, 1)

The distinction between the language under investigation—the *object language*—and the language in which the investigation is carried out—the *metalanguage*—was indispensable for Tarski's method, and so it was for Carnap. They both knew the consequences of not respecting the distinction. Carnap's *syntax-language* could either be a natural word-language, or a symbol-language, or even a mixture of words and symbols. The essential difference between the languages Tarski and Carnap were interested in is the meaning of the symbols. Tarski states explicitly that

> ... we are not interested here in 'formal' languages and science in one special sense of the word 'formal', namely sciences to the signs and expressions of which no (intuitive) [cf. Tarski (1933, 33)] meaning is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful. (Tarski 2006, 166)

Such formal languages are precisely the languages which constitute Carnap's field of investigations. Influenced by Hilbert's methods of formalizing mathematical theories, hence carrying out the investigations at a metamathematical level, Carnap also refrained from assigning any kind of interpretation to the signs and expressions of the language. However, there are differences between Hilbert's program and Carnap's goals.

> Whereas Hilbert intended his metamathematics only for the special purpose of proving the consistency of a mathematical system formulated in the object-language, I aimed at the construction of a general theory of linguistic forms. (Carnap 1963, 54)

Although Carnap praised Hilbert's formal methods and adopted the view that formal expressions of the language possess no meaning, he departed from Hilbert's methodology. Carnap's syntax ranges not only over metamathematics, but also over other formalized languages, as a means to formalize science in general (cf. Wagner 2009, 15–16).

Another essential parallel is clearly visible between Tarski's application of the notion of *satisfaction* and Carnap's *evaluation* introduced in §34c. The notion of *valuation* plays a crucial role in Carnap's definition of the term "analytic", actually a very similar role to that played by the notion of *satisfaction* in Tarski's definition of truth. In fact, Carnap's definition of "analytic in Language II" can be understood, for certain languages, as a definition of "true in Language II". One possible reason why Carnap did not put forward a definition of truth, in spite of coming so close to defining truth in a manner very similar to Tarski's "σ_1 is true in S", is that it would require an exposition of the meanings of the symbols occurring in the sentence of which truth is predicated, and that would go beyond Carnap's syntactical method (cf. Wagner 2009, 26).

What is important for our present discussion is how or, if at all, Carnap made a "turn" from syntax to semantics. We lack any documented evidence on when Carnap exactly reached certain "semantic" results. He finished LSL in December of 1933, two years before the publication of the German version of CTFL. We know that he discussed his investigations with Tarski and Gödel (see Woleński 1999, 8) as Carnap himself remembers:

> Even before the publication of Tarski's article I had realized, chiefly in conversations with Tarski and Gödel, that there must be a mode, different from the syntactical one in which to speak about language. Since it is obviously admissible about facts, and, on the other hand, notwithstanding Wittgenstein, about expressions of a language, it cannot be inadmissible to do both in the same language. (Carnap 1963, 60)

The fact is that the results Carnap achieved and even the methods he applied—truth, the undefinability theorem, the definition of mathematical truth using evaluation, the definitions of "analytic" (true) and "contradictory" (false)—are all semantic concepts! It has often, rightly in my opinion, been claimed (e.g., by Wagner, Woleński and others) that Carnap's syntax was actually "semantics in disguise". There are obvious parallels between the methods employed by Tarski in CTFL and by Carnap in LSL. In the part on general syntax, Carnap outlines a general method of defining "true in S_1", where S_1 is an object language, in a metalanguage S_2, and makes a statement on semantical paradoxes.

> This contradiction only arises when the predicates 'true' and 'false' referring to sentences in a language S are used in S itself. On the other hand, it is possible to proceed without incurring any contradiction by employing the predicates 'true (in S_1)' and 'false (in S_1)' in a syntax of S_1 which is not formulated in S_1 itself but in another language S_2.. . . A theory of this kind formulated in the manner of a syntax would nevertheless not be genuine syntax. *For truth and falsehood are not proper syntactical properties;* whether a sentence is true or false cannot generally be seen by its design, that is to say, by the kinds and serial oder of its symbols. [This fact has usually been overlooked by logicians, because, for the most part, they have been dealing not with descriptive but only with logical languages, and in relation to these, certainly, 'true' and 'false' coincide with 'analytic' and 'contradictory', respectively, and are thus syntactical terms.] (Carnap 1937, 216)

If the truth or the falsehood of a sentence follows from the rules of transformation of the language in which the sentence is given then we can translate "true" by "valid" or "analytic" and "false" by 'contravalid' or "contradictory" (cf. Carnap 1937, 216–17). There is another complicated issue involving "analytic" which, however, goes beyond the scope of this paper (see Coffa 1987).

There is a respectable amount of literature (e.g., Coffa, Ricketts, Woleński, Patterson, Creath to name only a few) comparing Carnap's and Tarski's methods and results. In spite of almost opposite approaches and methods, their goals were very much alike, and they both profited greatly from each other's results. What is of crucial importance for the postscript of CTFL is Carnap's systems of levels presented in §53 of LSL. Carnap defined it as an ordered series of non-empty classes of expressions. He emphasized that since the number of the expressions of a language is denumerably infinite, so is the number of their classes. These classes were called *levels* and were to be numbered with finite, or, if necessary, also transfinite numbers (Carnap 1937, 186–89). Although Carnap's characterization of his system of levels manifests certain similarities to Tarski's theory of semantical categories and their orders, the major differences are of essential importance, e.g., Carnap's introduction of transfinite numbers. We shall return to Carnap's LSL in the next section to present the exact contribution of his system to Tarski's definition of truth.

Even before LSL was published, the book was read and commented on by outstanding logicians and philosophers, among them Gödel, Quine, and Schlick. Carnap emphasizes the influence other scholars had on LSL.

> The point of view of the formal theory of language (known as "syntax" in our terminology) was first developed for mathematics by Hilbert in his "meta-mathematics", to which the Polish logicians, especially Adjukiewicz, Leśniewski, Łukasiewicz and Tarski, have added a "meta-logic". For this theory, Gödel created his fruitful method of "arithmetization". On the standpoint and method of syntax, I have, in particular, derived valuable suggestions from conversations with Tarski and Gödel. (Carnap 1937, xvi (1934 Preface))

It is essential for our further argument that we can be certain that Tarski knew Carnap's monograph of 1934. In a post card to Twardowski, dated 10 May 1934, Tarski wrote that Carnap sent him a correction of his new book *Die logische Syntax der Sprache*; thereupon Tarski suggested the adoption of the terminology used by Carnap in LSL for the German translation of CTFL.[2] Tarski also suggested multiple corrections which appeared in the 1937 English edition of LSL. Carnap acknowledges Tarski's contribution in the Preface to the English edition.

> The majority of these corrections and a number of further ones

[2] A post card numbered L 149/34 archived by Polskie Towarzystwo Filozoficzne (Polish Philosophical Society) in Poznań.

> have been suggested by Dr. A. Tarski, others by J. C. C. McKinsey and W. V. Quine, to all of whom I am very much indebted for their most helpful criticisms. (Carnap 1937, xi)

An apt young scholar himself, Tarski understood quickly how valuable Carnap's theory of levels was for his definition of truth, and that it would enable him to reach positive results where Leśniewski's framework did not work.

4. Tarski 1935

Postscript

Influenced by his *Doktorvater*, Leśniewski, in the original version from 1933, Tarski committed himself to working within his interpretation of STT. This fact significantly influenced the entire work and its final results.

> It seemed to me then that 'the theory of the semantical categories penetrates so deeply into our fundamental intuitions regarding the meaningfulness of expressions, that it is hardly possible to imagine a scientific language whose sentences possess a clear intuitive meaning but whose structure cannot be brought into harmony with the theory in question in one of its formulations' (cf. p. 215). Today I can no longer defend decisively the view I then took of this question. (Tarski 2006, 268)

Two years later, after having read LSL, it seemed important for Tarski to also investigate the formalized languages for which the fundamental principles of the theory of semantical categories no longer hold. In the postscript Tarski abandoned STT and turned to a new framework, which has been interpreted by some (see Sundholm 2003, 119–20), not entirely correctly, to be set theory. Let's look at a brief outline of the new method presented by Tarski in the postscript. The languages now investigated exhibit in their structure the greatest possible analogy with the languages previously studied, except for the differences connected with the theory of semantical categories. Just as in §2 and §4 Tarski specifies the basic concepts for the newly investigated languages (primitive sentential function, axiom, consequence, provable theorem etc). Here, the concept of *order* of an expression, introduced in §4, also plays an essential part, but as we will shortly see, Tarski gives the reader a new perspective on this notion.

To the names of individuals and to the variables representing them Tarski assigns order 0 (and not 1 as before). This is a direct parallel between Tarski's framework and Carnap's theory of levels.

> By a **system of levels** in S, we understand an ordered series $\mathfrak{R}_1$ of non-empty classes of expressions which fulfil the six conditions given on p. 188. Since the number of the expressions of a language is, at the most, denumerably infinite, the number of classes of $\mathfrak{R}_1$ is likewise at the most denumerably infinite. These classes we call **levels**; let them be numbered with the finite–and, if necessary, also with the transfinite–ordinal numbers (of the second number-class): level 0 (or the zero level), level $1, 2, \ldots \omega, \omega + 1 \ldots$ We shall designate the expressions which belong to the classes of $\mathfrak{R}_1$ by '$\mathfrak{Stu}$' [*Stufe*]; and, specifically, those which belong to level α (where 'α' designates an ordinal number) by '${}^{\alpha}\mathfrak{Stu}$'. (Carnap 1937, 186–7)

For Carnap "The ${}^{0}\mathfrak{Stu}$ are called *individual expressions* and, as symbols, individual symbols" (Carnap 1937, 188).

The order of a sentence-forming functor of a sentential function has been previously unambiguously determined by the orders of all arguments of this function, but now the principles of STT no longer apply. The theory of levels allows both expressions of infinite order and predicates and functors that take arguments of variable order. For Tarski, *orders* could all be numbered by finite or transfinite ordinal numbers, just as Carnap's levels. The fact that the level of the arguments of a predicate is not fixed, but variable, allows Tarski to introduce the variables which "run through" all orders. Therefore,

> . . . it may happen that one and the same sign plays the part of

> a functor in two or more sentential functions in which arguments occupying respectively the same places nevertheless belong to different orders. Thus in order to fix the order of any sign we must take into account the orders of all arguments in all sentential functions in which this sign is a sentence-forming functor. (Tarski 2006, 269)

In order to classify the signs of infinite order, Tarski employs the notion of *ordinal number*, taken from set theory, which is a generalization of the concept of natural number—the smallest ordinal numbers. For every infinite sequence of ordinal numbers there are numbers greater than every term of the sequence, there are also numbers which are greater than all natural numbers. These are *transfinite ordinal numbers*. In every non-empty class of ordinal numbers there is the smallest ordinal number, hence also the smallest transfinite number—denoted by the symbol "ω". To the signs of infinite order which are functors of sentential functions containing exclusively arguments of finite order we assign the number "ω" as their order (e.g. the language of the general theory of classes has the order ω). These explications are followed by a general recursive definition of order used by Tarski: "the order of a particular sign is the smallest ordinal number which is greater than the orders of all arguments in all sentential functions in which the given sign occurs as a sentence-forming functor.[[2]]" The footnote [[2]] takes us directly to the introduction of the system of levels in Carnap's LSL. Tarski realized that in order to define truth for "superior" languages, it was crucial that the variables in the languages investigated now were not of a definite order.

> ... we must introduce into the languages variables of indefinite order which, so to speak, 'run through' all possible orders, which can occur as functors or arguments in sentential functions without regard to the order of the remaining signs occurring in these functions, and which at the same time may be both functors and arguments in the same sentential functions. (Tarski 2006, 271)

This is precisely the reason why Tarski is now able to define truth for the languages of infinite order. By admitting the expressions of transfinite order, Tarski allows for variables to be of indefinite order, which in turn means that variables can act as functors or arguments in sentential functions, or even in the same sentential function, at the same time disregarding the order of other signs in this function. As Patterson (2012, 172) notes, using expressions of infinite order is not merely a matter of adding transfinite levels atop the hierarchy of STT, and thus being able to define truth for the general theory of classes in a languages which adheres to the principles of semantical category. Referring to Sundholm (2003, 118), Patterson concludes the following:

> Since Tarski wants to adhere to the principle that the order of an expression is the least ordinal greater than any that specifies the order of any argument it takes, but there is no finite ordinal α such that ω is the least ordinal greater than α, the only way to get expressions of transfinite order is to have expressions that take arguments of *all* finite orders and hence to allow for variability in the order of the arguments that a functional expressions takes. (Patterson 2012, 172)

Following these elucidations is the often quoted footnote in which Tarski explains the notion of order for the languages considered in this article. It has been argued (e.g., by Patterson 2012, 172) that Tarski's change of logical framework causes an ambiguity regarding the notion of order. The ambiguity, however, occurs only if we apply the method presented in the postscript to the languages of set theory. Since Tarski was not working with set theory but with Carnap's theory of levels, there is no ambiguity in the notion of order; both Tarski and Carnap apply the notion of order to expressions of the language, hence it is a syntactical notion in both cases.[3] In this

[3] For a detailed discussion on this topic see Loeb (2014) and de Rouilhan (1998). Loeb also presents an interesting argument regarding Tarski's change of logical framework.

context, it is important to notice that from the languages Tarski considers in the postscript it is *but a step* to languages of another kind. The languages of another kind are the languages of set theory, such as presented by Zermelo and his successors.

> For the languages here discussed the concept of order by no means loses its importance; it no longer applies, however, to expressions of the language, but either to the objects denoted by them or to the language as a whole. Individuals, i.e. objects which are not sets, we call objects of order 0; the order of an arbitrary set is the smallest ordinal number which is greater than the orders of all elements of this set; the order of the language is the smallest ordinal number which exceeds the order of all sets whose existence follows from the axioms adopted in the language. Our further exposition also applies without restriction to the languages which have just been discussed. (Tarski 2006, 271n)

Even though the postscript itself is not written within set theory, Tarski emphasizes in the last sentence of this footnote that his expositions apply without restriction also to the languages of set theory. Tarski changes the logical framework from his interpretation of type theory, influenced by Leśniewski, to Carnap's theory of levels. In the language he investigates, following Carnap's lead, he assigns order 0 to names of individuals and to variables representing them, not 1 as before.

The essential move is the introduction of variables of transfinite order not only to the investigated (object) language, but also to the metalanguage in which the investigations are carried out. This allows for the metalanguage to be constructed in such a way that it contains *variables of higher order* than the variables of the object language and thus, to become an *essentially richer* language. It is precisely the essential richness of the metalaguage which constitutes it as a language of higher order than the object language. In STT, in which Tarski was working in the Polish original, the order of each category determined the orders of all expressions belonging to this category, i.e. all expressions belonging to a given semantical category had the same order assigned to them—called the *order of this category* (Tarski 2006, 218). However, the theory of semantical categories worked only within the languages of finite order. In the postscript Tarski turns to Carnap's system of levels and thus, changes the logical framework to allow for expressions to be of transfinite order, and more importantly for expressions which do not determine the orders of their arguments (Tarski 2006, 270n). Ray (2005) presents an argument for an interpretation of the notion of order as used by Tarski. Ray notes that

> . . . a language might be of a higher order for either of two distinct reasons. In its original formulation the only way to have a *language* of higher order was to have *variables* of higher order. Call this limited notion *higher order in the narrow sense*. However, as a result of this extension of the notion of order to languages like the language of Zermelo set theory, it becomes possible to have a language of higher order but which does *not* have variables of higher type (nor any difference of grammatical form at all). This is because the order of the language in these cases is determined by the "order of all sets whose existence follows from the axioms adopted in the language".[[2]] Thus, under some circumstances the order of a language could be increased merely by the addition of an axiom. Call the notion which allows for this *higher order in the extended sense*. (Ray 2005, 436)

The definition of higher order languages in the extended sense, as presented by Ray, applies to the languages of set theory, e.g. the language of Zermelo set theory. For the language of Carnap's theory of levels, according to Ray's distinction, the notion of higher order in the narrow sense applies. Further, Ray points us in the direction of Tarski's 1944 "simplified" version of the manuscript, where Tarski upheld this informal definition of *essential richness*. He held that in the construction of the required definition of truth using the recursive definition of satisfaction we need to

> . . . introduce into the meta-language variables of a higher logical type than those which occur in the object-language; or else to as-

> sume axiomatically in the meta-language the existence of classes that are more comprehensive than all those whose existence can be established in the object-language. (Tarski 1944, 353n)

The second condition applies to the languages of set theory. Working within Carnap's theory of levels, Tarski emphasizes that we can always introduce into the metalanguage variables of higher order than all the variables of the object language. This means that the metalanguage can always be constructed in such a way to become a language of higher order than the object language.

> In particular it is always possible to construct the metalanguage in such a way that it contains variables of higher order than all the variables of the language studied. The metalanguage then becomes the language of higher order and thus one which is essentially richer in grammatical forms than the language we are investigating. This is a fact of the greatest importance from the point of view of the problems in which we are interested. For with this the distinction between languages of finite and infinite orders disappears—a distinction which was so prominent in §§ 4 and 5 and was strongly expressed in the theses A and B formulated in the *Summary*. (Tarski 2006, 271–72)

This means that a construction of a formally correct and materially adequate definition of *true sentence* for languages of infinite order is now possible, as long as the metalanguage is of higher order than the object language. Tarski makes this statement explicit in the new theses.

> A. *For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language—but under the condition that the metalanguage possesses a higher order than the language which is the objet of investigations.*
>
> B. *If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed.* (Tarski 2006, 273)

When we compare the new theses with the old ones, we notice at once that Tarski went down from three statements to only two. The original statement C loses its importance in light of thesis A. The newly written thesis A states clearly that a formally correct and materially adequate definition of *true sentence* can be constructed for *every*— finite or infinite—formalized language as long as the metalanguage is of higher order than the object language. With this statement Tarski rewrote the final results of his original paper. At the same time Tarski emphasizes that the results presented in Theorem I of §5 are still valid and can be extended to languages of any order. Theorem I, which has often later been called the "undefinability theorem" (see e.g., Field 2008, 27), states that it is impossible to give an adequate definition of truth for a language in which the order of the metalanguage does not exceed the order of the investigated language.

In defining truth for the languages of indefinite order the essential step is the introduction of variables of transfinite order not only to the object language but also to the metalanguage. This allows for the construction of a higher order metalanguage which is essentially richer in grammatical forms than the language studied. This step cancels the distinction between the languages of finite and infinite order which yielded the negative conclusion in §5 of the original paper. Thus, as Tarski notes in retrospect

> . . . the setting up of a correct definition of truth for languages of infinite order would in principle be possible provided we had at our disposal in the metalanguage expressions of higher order than all the variables of the language investigated. The absence of such expressions in the metalanguage has rendered the extension of these methods of construction to languages of infinite order impossible.

> But now we are in a position to define the concept of truth for any language of finite or transfinite order, provided we take as the basis for our investigations a metalanguage of an order which is at least greater by 1 than that of the language studied (an essential part is played here by the presence of variables of indefinite order in the metalanguage) (Tarski 2006, 272)

Our short explication of the parallels between Tarski's CTFL and Carnap's LSL highlights the reasons motivating Tarski's move from Leśniewski's version of STT to Carnap's theory of levels, and thus the reasons for writing the postscript.

5. Conclusion

In the original Polish version of CTFL, Tarski showed how to construct a definition of *true sentence* for the languages of finite order. The semantical categories, or the unifying category, of the metalanguage could not be of lower order than any one of the variables of the object language. In the postscript Tarski no longer subscribed to Leśniewski's version of type theory. Instead, he worked with Carnap's theory of levels. The essential feature of this theory is that it employs expressions of transfinite order, which in turn, allows for the variables to be of indefinite order. Now the variables could "run through" all possible orders, which is exactly what Tarski needed to define truth for languages of infinite order. By means of Carnap's theory of levels, Tarski was able to modify his original conclusions, and thus to state that it is always possible to construct a definition of *true sentence* provided we have at our disposal a metalanguage possessing expressions of higher order than all the variables of the object language. With this statement Tarski rewrote the final results of CTFL. Tarski sustains the results presented here in the later simplified version of his monumental work:

> It turns out, however, that this [axiomatic] procedure can be avoided. For *the condition of the "essential richness" of the meta-language proves to be, not only necessary, but also sufficient for the construction of a satisfactory definition of truth;* i.e., if the meta-language satisfies this condition, the notion of truth can be defined in it. (Tarski 1944, 351)

Tarski's abandonment of the logical system adopted by his *Doktorvater*, Leśniewski, in favor of Carnap's theory of levels is a crucial step on Tarski's philosophical path.[4] It is a step which Tarski upheld in his later works, e.g., "On the Concept of Logical Consequence", where he referred to Carnap's LSL and praised him for the construction of a first precise definition of the concept of consequence. As we have seen, when closely examined, the parallels between Tarski's CTFL and Carnap's LSL are indisputable. It should be remembered that the concepts "semantic" and "syntax" went through a revolutionary period. With Tarski's CTFL and Carnap's LSL the meanings of these terms changed significantly, especially for Carnap, but also for every philosopher and logician, henceforth.

> Although Carnap, Gödel and Tarski did not use the terminology stemming from the arithmetical hierarchy, they contributed to the effect that semantics is essentially stronger than syntax. The distribution of merits is difficult and in fact secondary. The most important point is that semantics provides methods which give an opportunity for finite minds to deal with infinity. It is not strange that these methods have to be non-finitary. Judging the general philosophical significance of semantics is still far from being finished. (Woleński 1999, 12)

It was important for our argument to establish that Tarski had known Carnap's LSL (1934) as he wrote the postscript (1935). One could naturally argue that it is possible that Tarski arrived at the idea of transfinite types independently of Carnaps's work on levels, for example through his own work on set theory.

[4] This issue has been thoroughly discussed by Tarski's specialists in numerous publications, e.g., Betti, Feferman, Sundholm, Woleński to mention only a few.

Even though this is not utterly impossible, Tarski would most likely have made an explicit statement on this, just as he did in *Historical Notes* and in a footnote on page 247 in regard to his and Gödel's results on the indefinability of truth. Tarski makes no such statement about Carnap's theory of levels. Considering the historical facts and the content of *Postscript*, an option that Tarski did in fact decide to use Carnap's theory of levels after having read his LSL must be allowed for. Both Carnap and Tarski acknowledged each other's work and felt indebted for the contribution it made to their own future research. While we can observe Carnap's slow turn from the strictly syntactical method towards a semantic one, we know that he never gave up his *Principle of Tolerance*. In turn, Tarski's recognition of Carnap's work can be seen as his move towards Carnap's *Principle of Tolerance*, and thus agreeing that there are more logical systems which are equally acceptable.

Acknowledgements

I'd like to thank Hartry Field for pointing out the controversy of the Postscript to me and for the following discussions. My special thanks go to Iris Loeb who discussed this topic with me over the years and with whom, contrary to our original discussions, I actually agree on some points in the end. I thank Leon Horsten for discussion on sets and classes and Hannes Leitgeb for his very helpful and specific remarks. I'd also like to thank Christian Damböck for inviting me to *Philosophy of Science Colloquium – Institute Vienna Circle* and for his questions and naturally, thanks to all participants for their discussion. Last but not least, I especially thank Richard Zach, Kevin C. Klement and two anonymous referees for their most helpful comments and suggestions.

Monika Gruber
University of Vienna
monika.gruber@univie.ac.at

References

Awodey, S. and A. W. Carus, 2009. "From Wittgenstein's Prison to the Boundless Ocean: Carnap's Dream of Logical Syntax." In Wagner (2009), pp. 79–106.

Betti, A., 2008. "Polish Axiomatics and its Truth: On Tarski's Leśniewskian Background and the Adjukiewicz Connection." In *New Essays on Tarski and Philosophy*, edited by D. Patterson, pp. 44–72. Oxford: Oxford University Press.

Carnap, R., 1934. *Logische Syntax der Sprache*. Vienna: Springer. Translated as Carnap (1937).

———, 1937. *The Logical Syntax of Language*. London: Kegan Paul, Trench, Trubner & Co. Ltd. A translation of Carnap (1934) by A. Smeaton.

———, 1963. "Intellectual Autobiography." In *The Philosophy of Rudolf Carnap*, edited by P. A. Schlipp, pp. 3–84. La Salle, IL: Open Court.

Coffa, A., 1987. "Carnap, Tarski and the Search for Truth." *Noûs* 21: 547–572.

Coquand, T., 2010. "Type Theory." *Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/entries/type-theory/` (accessed February 20, 2014).

Creath, R., 1999. "Carnap's Move to Semantics: Gains and Losses." In Woleński and Köhler (1999), pp. 65–76.

Davidson, D., 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

de Rouilhan, P., 1998. "Tarski et l'université de la logique: Remarques sur le post-scriptum au 'Wahrheitsbegriff'." In *Le formalisme en question: Le tournant des annés 30*, edited by F. Nef and D. Vernant, pp. 85–102. Paris: Vrin, Problèmes et controverses.

Feferman, S., 2002. "Conceptual Analysis of Semantical Notions." *Stanford Encyclopedia of Philosophy*. `http://math.stanford.edu/~feferman/papers/conceptanalysis.pdf` (accessed April 20, 2014).

Ferreirós, J., 2007. *Labyrinth of Thought. A History of Set Theory and Its Role in Modern Mathematics*. Basel: Birkhäuser Verlag AG, second revised ed.

Field, H., 2008. *Saving Truth from Paradox*. Oxford: Oxford University Press.

Friedman, M., 1999. *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

Frost-Arnold, G., 2004. "Was Tarski's Theory of Truth Motivated by Physicalism?" *History and Philosophy of Logic* 25: 265–280.

Gómez-Torrente, M., 2004. "The Indefinability of Truth in the 'Wahrheitsbegriff'." *Annals of Pure and Applied Logic* 126: 27–34.

———, 2011. "Alfred Tarski." *Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/entries/tarski/` (accessed February 20, 2014).

Gruber, M., 2013. "Tarski, Blow by Blow." PhD Thesis, University of Salzburg.

Gupta, A. and N. Belnap, 1993. *The Revision Theory of Truth*. Cambridge, MA: MIT Press.

Horwich, P., 1998. *Truth*. Oxford: Oxford University Press.

Jadacki, J., ed., 2003. *Alfred Tarski: dedukcja i semantyka (déduction et sémantique)*. Warsaw: Wydawnictwo Naukowe *Semper*.

Kripke, S., 1975. "Outline of a Theory of Truth." *The Journal of Philosophy* 72: 690–716.

Leśniewski, S., 1929. "Grundzüge eines neuen Systems der Grundlagen der Mathematik." *Fundamenta Mathematicae* 14: 1–81.

Loeb, I., 2014. "Towards Transfinite Type Theory: Rereading Tarski's *Wahrheitsbegriff*." *Synthese* 191: 2281–2299.

Mormann, T., 2000. *Rudolf Carnap*. Munich: Verlag.

Nowaczyk, A., 2003. "Co naprawdę powiedział Tarski o prawdzie w roku 1933?" In Jadacki (2003), pp. 61–76.

Patterson, D., 2012. *Alfred Tarski: Philosophy of Language and Logic*. Basingstoke: Palgrave Macmillan.

Quine, W. V., 1963. *Set Theory and Its Logic*. Cambridge, MA: Harvard University Press.

Ray, G., 2005. "On the Matter of Essential Richness." *Journal of Philosophical Logic* 34: 433–457.

Schurz, G., 1999. "Tarski and Carnap on Logical Truth – or: What Is Genuine Logic?" In Woleński and Köhler (1999), pp. 77–94.

Sundholm, B. G., 2003. "Tarski and Leśniewski on Languages with Meaning versus Languages without Use: A 60th Birthday Provocation for Jan Woleński." In *Philosophy and Logic: In Search of the Polish Tradition*, edited by J. Hintikka, T. Czarnecki, K. Kijania-Packet, T. Placek, and A. Rojszczak, pp. 109–128. Dordrecht: Kluwer.

Tarski, A., 1933. "Pojęcie prawdy w językach nauk dedukcyjnych." Reprinted in Zygmunt, J., ed., *Tarski, A. Pisma logiczno-filozoficzne. Tom 1. Prawda.* Warsaw: Wydawnictwo Naukowe PWN, 1995, pp. 3–172.

———, 1935. "Der Wahrheitsbegriff in den formalisierten Sprachen." *Studia Philosophica* 1: 261–405.

———, 1936. "The Concept of Logical Consequence." Reprinted in *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, 2nd ed. Indianapolis, Hackett Publishing Company, 2006, pp. 408–420.

———, 1944. "The Semantic Conception of Truth and the Foundations of Semantics." *Philosophy and Phenomenological Research* 4: 341–376.

———, 2006. "The Concept of Truth in Formalized Languages." Reprinted in *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, 2nd ed. Indianapolis, Hackett Publishing Company, 2006, pp. 152–278.

Uebel, T., 2009. "Carnap's *Logical Syntax* in the Context of the Vienna Circle." In Wagner (2009), pp. 53–76.

Wagner, P., ed., 2009. *Carnap's Logical Syntax of Language*. Basingstoke: Palgrave Macmillan.

Woleński, Jan, 1999. "Semantic Revolution – Rudolf Carnap, Kurt Gödel, Alfred Tarski." In Woleński and Köhler (1999), pp. 1–16.

———, 2003. "Języki sformalizowane a prawda." In Jadacki (2003), pp. 67–76.

Woleński, Jan and Eckehart Köhler, eds., 1999. *Alfred Tarski and the Vienna Circle: Austro-Polish Connections in Logical Empiricism*. Dordrecht: Kluwer Academic Publishers.