THE BULLETIN OF SYMBOLIC LOGIC
Volume 19, Number 4, Dec. 2013

LOGIC IN THE 1930s: TYPE THEORY AND MODEL THEORY

GEORG SCHIEMER AND ERICH H. RECK

Abstract. In historical discussions of twentieth-century logic, it is typically assumed that model theory emerged within the tradition that adopted first-order logic as the standard framework. Work within the type-theoretic tradition, in the style of *Principia Mathematica*, tends to be downplayed or ignored in this connection. Indeed, the shift from type theory to first-order logic is sometimes seen as involving a radical break that first made possible the rise of modern model theory. While comparing several early attempts to develop the semantics of axiomatic theories in the 1930s, by two proponents of the type-theoretic tradition (Carnap and Tarski) and two proponents of the first-order tradition (Gödel and Hilbert), we argue that, instead, the move from type theory to first-order logic is better understood as a gradual transformation, and further, that the contributions to semantics made in the type-theoretic tradition should be seen as central to the evolution of model theory.

§1. Introduction. The development of mathematical logic in the past century is characterized by the gradual separation of syntax and semantics and by the corresponding consolidation of two metalogical research programs, proof theory and model theory. While proof theory took shape in the 1920s and '30s, mainly in the Hilbert school, model theory as practiced today was established as late as the 1950s, largely as a result of work by Alfred Tarski.[1] In recent studies of the emergence of model theory, the main focus has been on two aspects: early contributions within the algebraic tradition in logic, exemplified by writings of Schröder, Löwenheim, and Skolem; and the "semantic revolution" in the 1930s, in particular Tarski's publications on truth and logical consequence.[2] In the present paper, we will provide a novel perspective on the latter, or more generally, on the history of model theory in the 1930s.

Received February 29, 2012.

[1]For model theory, see Tarski [1954] and Tarski and Vaught [1957]. Formal models were first defined in a series of papers by John Kemeny, especially Kemeny [1956a] and Kemeny [1956b]. See Hodges [2009] for the development of model theory after 1945.

[2]See Badesa [2004] for the algebraic tradition in logic, especially Löwenheim's work. See Mancosu [2010a] for an overview of work on Tarski's early contributions.

1079-8986/13/1904-0001/$5.00
DOI:10.2178/bsl.1904010

Historical discussions of twentieth-century logic tend to be shaped by the distinction between "two logical frameworks"—type theory and first-order logic—and the view that different, partly incompatible conceptions of logic are associated with them.[3] It is also typically assumed that model theory emerged in the tradition that adopted first-order logic as the standard logical system. In contrast, type theory in the tradition of Russell and Whitehead's *Principia Mathematica* tends not to be seen as an adequate framework for model-theoretic semantics. Work within the type-theoretic tradition is thus not associated with the rise of model theory, despite the fact that simple type theory (henceforth STT) was a widely used, or indeed the main, logical framework in the 1930s, including for early work in semantics.[4] As a consequence, the transition from semantics in type theory to a model-theoretic approach based on first-order languages has not been studied in detail yet. More generally, the transition from logic in the style of *Principia Mathematica*, into the 1930s, to "modern" model-theoretic logic, from the 1950s on, remains one of the greatly understudied parts of the history of logic.

In the present paper, our main goal is to begin filling this gap. For that purpose, we will provide a survey of formative work in semantics within STT during the 1930s. Specifically, we will consider several attempts by Carnap and Tarski to express the semantics of axiomatic theories within STT prior or parallel to the turn to the now standard metatheoretic semantics. These attempts will also be compared to work from the same period by two proponents of the first-order tradition: Gödel and Hilbert. By comparing their definitions of "satisfaction", "truth in a model", and "validity", we will illustrate that the two logical traditions are less far apart on semantic issues than is usually believed. In fact, we will try to establish—as our second main goal—that the shift from type theory to model theory is better understood as a gradual transformation than as a radical break, and further, that the contributions to semantics in the 1930s made within the type-theoretic tradition have to be seen as central to this evolution.

The main body of this paper is divided into three sections. In Section 2, we present now standard definitions of several key model-theoretic notions, as background for our later discussions (Section 2.1); we also review ways in which the semantics of a simple type-theoretic language can be represented within the same language in terms of a hierarchy of partial truth predicates (Section 2.2). In Section 3, a variety of historical attempts to define semantic notions—by Carnap, Tarski, Gödel, and Hilbert—will be surveyed (Sections 3.1–3.6), including highlighting similarities and other

[3] See, e.g., Goldfarb [1979] and Hintikka [1988]. For further discussions of the two logical traditions, see Ferreirós [2001] and Mancosu [2010a].

[4] Ferreirós [2007] gives a general survey of the applications of STT in the 1930s. We will discuss corresponding contributions to formal semantics in detail below.

connections between them (Section 3.7). In Section 4, we compare these earlier approaches further with their current counterparts, thus analyzing the ways in which, or the extend to which, the former anticipated the latter. In that connection, two issues receive special attention: In Section 4.1, we juxtapose two ways of formulating model theory—metatheoretically and "monolinguistically"—that coexisted in work from the 1930s; and in Section 4.2, we address the topic of domain variability, i.e., the question to what extent, if any, these early treatments of model-theoretic truth allow for a variation of model domains. The paper ends with a brief summary and conclusion.

§**2. Truth in a model.** The main focus in this paper will be on several definitions of *satisfaction*, *truth*, and *general validity* (or *logical truth*) that were articulated by Carnap, Gödel, Tarski, and Hilbert, respectively, in the 1930s. But before turning to their original works, a brief presentation of these notions in their current form will be provided, thus also fixing notation. This will set the stage for a closer comparison between the old and the new approaches.

2.1. Modern metatheoretic definitions. Let $\mathcal{L}$ be a first-order language, containing logical symbols for negation, conjunction, and the existential quantifier,[5] a denumerable number of individual variables, and, as non-logical vocabulary, a denumerable number of individual constants, n-ary predicates, and n-ary functions (for every n). Terms and formulas of $\mathcal{L}$ can then be defined recursively, as usual.

Now let the structure $\mathcal{M} = (D, I)$ be an *interpretation* of $\mathcal{L}$, where D is a non-empty universe and I an interpretation function that assigns elements of D to the individual constants, relations on D to the predicates, and functions on D to the function symbols. (Each n-ary predicate is assigned a subset of D^n, each n-ary function symbol a function $f\colon D^n \to D$.) Instead of interpretations we will also talk about *models* in this connection. An assignment s for the model $\mathcal{M}$ is then a function from the set of variables V of $\mathcal{L}$ to D, i.e., $s\colon V \to D$. The range of s can be considered as an ordered sequence of elements of D, of the form $\langle o_1, o_2, o_3, \dots \rangle$, where each o_i is assigned to a specific variable x_i of $\mathcal{L}$. (Instead, one can restrict oneself to arbitrarily long finite sequences $\langle o_1, o_2, \dots, o_n \rangle$ for every n. In some contexts this has technical advantages, which will play a role later on.) Relative to a model $\mathcal{M}$ and a corresponding assignment s, the *denotation* of terms t of $\mathcal{L}$, designated by $\|t\|_s^{\mathcal{M}}$, is again defined recursively: (i) if t is a variable v, then $\|v\|_s^{\mathcal{M}} = s(v)$; (ii) if t is an individual constant c, then $\|c\|_s^{\mathcal{M}} = I(c)$; (iii) if t is a functional expression of the form $f(t_1, \dots, t_n)$, where $t_1, \dots, t_n$ are terms, then $\|f(t_1, \dots, t_n)\|_s^{\mathcal{M}} = I(f)(\|t_1\|_s^{\mathcal{M}}, \dots, \|t_n\|_s^{\mathcal{M}})$.

[5]The other logical constants are assumed to be reduced to these, as usual.

On that basis, the notion of *satisfaction* of a formula φ in a model $\mathcal{M}$ relative to an assignment s, in symbols $\mathcal{M} \models \varphi[s]$, can be defined by recursion on the complexity of formulas, as follows:

1. If φ is an atomic formula of the form $R(t_1, \dots, t_n)$, then
$$\mathcal{M} \models \varphi[s] \text{ iff } \langle \|t_1\|_s^{\mathcal{M}}, \dots, \|t_n\|_s^{\mathcal{M}} \rangle \in I(R).$$
2. If φ is a negated formula of the form $\neg\psi$, then
$$\mathcal{M} \models \varphi[s] \text{ iff not } \mathcal{M} \models \psi[s].$$
3. If φ is a conjunctive formula of the form $\psi \wedge \chi$, then
$$\mathcal{M} \models \varphi[s] \text{ iff } \mathcal{M} \models \psi[s] \text{ and } \mathcal{M} \models \chi[s].$$
4. If φ is an existentially quantified formula of the form $\exists x\psi$, then
$$\mathcal{M} \models \varphi[s] \text{ iff for at least one assignment } s', \mathcal{M} \models \psi[s'].$$

In 4, the assignment s' differs from s at most in the value assigned to the free variable x. We call such assignments x-*variants* of s.[6]

When considering higher-order languages, the corresponding definitions of term, formula, model, assignment, and satisfaction are simple extensions of those for first-order languages. For instance, let $\mathcal{L}_2$ be a second-order extension of $\mathcal{L}$ such that, in addition to the vocabulary of $\mathcal{L}$, $\mathcal{L}_2$ contains a denumerable number of relation variables (of every arity n).[7] In this context, the formation rules allow for formulas with such variables bound by higher-order quantifiers. Let $\mathcal{M}' = (D, \{Rel^n\}_n, I)$ be a standard (or full) model of $\mathcal{L}_2$, where D and I are specified as before and $Rel^n = \wp(D^n)$. A second-order assignment function s for $\mathcal{M}'$ now also assigns elements of Rel^n to the n-ary relation variables of the language. For the definition of satisfaction, we need to add two corresponding clauses:

5. If φ is an atomic formula of the form $X(t_1, \dots, t_n)$, then
$$\mathcal{M}' \models \varphi[s] \text{ iff } \langle \|t_1\|_s^{\mathcal{M}'}, \dots, \|t_n\|_s^{\mathcal{M}'} \rangle \in s(X).$$
6. If φ is a formula of the form $\exists X\psi$, then
$$\mathcal{M} \models \varphi[s] \text{ iff for for at least one assignment } s', \mathcal{M}' \models \psi[s'].$$

In 6, s' differs from s again at most in the value for the variable X, i.e., it is an X-variant of s.[8]

[6] Instead of working with assignment functions in this context, one can also employ additional metatheoretic constants that name all the elements in a model. Indeed, since Shoenfield [1967] this may be the standard approach in model theory. We use the older (Tarskian) technique because it will be directly relevant for our historical case studies.

[7] Second-order function variables, for every arity n, can be added accordingly; or one can simulate their use by means of $n + 1$-place predicate variables, as usual.

[8] See Shapiro [1991, 70–72] for further details in the case of second-order languages.

Against that background, the following metalogical notions can be defined for formulas of both first and second-order languages (and, by suitable further extensions, for higher orders as well):

DEFINITION 1. φ *is* satisfiable *if there exist a model* $\mathcal{M}$ *and a corresponding assignment* s *such that* $\mathcal{M} \models \varphi[s]$.

DEFINITION 2. φ *is* true in a model $\mathcal{M}$, *in symbols* $\mathcal{M} \models \varphi$, *if it is satisfied under every assignment* s *for* $\mathcal{M}$.

DEFINITION 3. φ *is* valid, *or* logically true, *if it is true in every model of the language, i.e., if for all* $\mathcal{M}$: $\mathcal{M} \models \varphi$.

DEFINITION 4. $\mathcal{M}$ *is a model of a (finitely axiomatizable) theory* T *expressed in* $\mathcal{L}$, *in symbols* $\mathcal{M} \models T$, *if every sentence of* T *is true in* $\mathcal{M}$.

Our definition of satisfaction above was a standard recursive definition in the metatheory. As noted originally in Tarski [1935], it can be transformed into an explicit definition (using a general technique derived from Dedekind and Frege). That is to say, we can introduce a ternary satisfaction predicate *Sat* into the metalanguage such that

$$Sat(\varphi, \mathcal{M}, s) \text{ iff } \mathcal{M} \models \varphi[s]$$

The definition of this predicate has the following form: $Sat(\varphi, \mathcal{M}, s) \Leftrightarrow_{df}$ for all binary relations S, if $\Phi(\mathcal{M}, S)$ then $S(\varphi, s)$. Here Φ is defined as follows: $\Phi(\mathcal{M}, S) \Leftrightarrow_{df}$ for all formulas ϕ and and all assignments s for $\mathcal{M}$, $S(\phi, s)$ iff either (i) ϕ is an atomic formula of the form $R(t_1, \ldots, t_n)$ and $\langle \|t_1\|_s^{\mathcal{M}}, \ldots, \|t_n\|_s^{\mathcal{M}} \rangle \in I(R)$ or (ii) ϕ is a negated formula of the form $\neg\psi$ and not $S(\psi, s)$ or (iii) ϕ is a conjunctive formula of the form $\psi \wedge \chi$ and both $S(\psi, s)$ and $S(\chi, s)$ or (iv) ϕ is an existentially quantified formula of the form $\exists x\psi$ and there exists an assignment s' for $\mathcal{M}$ that is an x-variant of s such that $S(\psi, s')$.

Three further observation can be added. First, both the recursive and the explicit definition given so far are formulated in an informal metalanguage, i.e., colloquial English augmented by logical vocabulary. But they can be reformulated in a formalized metatheory, like ZFC, as long as certain standard techniques are available (including arithmetization of syntax). In particular, we can formalize the explicit definition so as to introduce a ternary satisfaction predicate Sat^* into the set-theoretic metalanguage $\mathcal{L}_\in$ such that for all formulas φ, all models $\mathcal{M}$, and all assignments s for $\mathcal{M}$

$$\text{ZFC} \vdash Sat^*(\ulcorner\varphi\urcorner, \mathcal{M}, s) \text{ iff } \mathcal{M} \models \varphi[s]$$

Here $\ulcorner\varphi\urcorner$ is the Gödel number of φ, $\mathcal{M}$ and s are treated as elements of V (assuming $\mathcal{M}$ is a set), and n-place relations as well as n-place functions are implemented as corresponding sets of tuples.[9] Second, such explicit

[9]The definition of Sat^* in $\mathcal{L}_\in$ will now be as follows: $Sat^*(\ulcorner\varphi\urcorner, s) \Leftrightarrow_{df} \forall A\, (\Phi(\mathcal{M}, A) \rightarrow \langle\ulcorner\varphi\urcorner, s\rangle \in A)$. The definition of Φ will have this form: $\Phi(\mathcal{M}, A) \Leftrightarrow_{df}$ for all Gödel numbers of sentences $\phi \in \mathcal{L}$ and all assignments s for $\mathcal{M}$, $\langle\ulcorner\phi\urcorner, s\rangle \in A$ iff either (i) $\ulcorner\phi\urcorner = \ulcorner R(t_1, \ldots, t_n)\urcorner$

definitions of satisfaction predicates can, once more, be extended to higher-order languages in a straightforward manner.[10] Third, and parallel to using $\mathcal{L}_{\in}$ as the metalanguage, formalized definitions of satisfaction for a language $\mathcal{L}$ can also be given in a suitable higher-order language $\mathcal{L}'$, as will be spelled out further in the next subsection.

2.2. Truth in type theory. As is well known, since Tarski's indefinability theorem, an adequate definition of semantic notions like satisfaction and truth can only be given in a metalanguage that is "essentially richer" than the object language. The necessary resources of the metalanguage with respect to the first-order language $\mathcal{L}$ became explicit in the definitions above. Namely, what needs to be expressible in the metalanguage is quantification over arbitrary assignments, or arbitrary sequences of objects of $\mathcal{M}$, in the satisfaction clause for quantified formulas. If we move on to an explicit definition, it additionally requires quantification over arbitrary relations between (Gödel numbers of) formulas and sequences of objects.[11] Thus, satisfaction and truth for formulas of a given language $\mathcal{L}$ cannot be directly represented in $\mathcal{L}$ itself. Tarski's theorem shows that they cannot be indirectly represented either.

At the same time, certain *partial* definability results for semantic notions within one and the same language are not ruled out. In the first-order case, partial truth predicates can be defined for fragments of a language by taking into account the quantificational complexity of formulas.[12] A second option—especially relevant for our purposes—is that of a "stratified theory of truth" for higher-order or type-theoretic languages. The idea here is to present a hierarchy of truth predicates for formulas of the same language but always of lower orders. The language can then express its own semantics, but not in terms of a single satisfaction predicate for all formulas. Rather, the higher levels are used to define typed "semantical constants" that apply to formulas from lower levels. (This idea was explored in work on semantics within the type-theoretic context during the 1930s, as we will see.) A formalized account of such a theory for a *ramified* type theory (RTT) was first presented systematically in Church [1976]. Church's approach cannot be discussed in any detail here. Instead, we will sketch a stratified theory of semantic predicates for a type-theoretic language that differs from his in two respects: (i) instead of RTT, we consider STT; (ii) instead of his intensional predicates *val*, we focus on the extensional predicates *Sat*.

---

and $\langle \|t_1\|_s^{\mathcal{M}}, \ldots, \|t_n\|_s^{\mathcal{M}} \rangle \in I(R)$ or (ii) $\ulcorner\phi\urcorner = \ulcorner\neg\psi\urcorner$ and $\langle \ulcorner\psi\urcorner, s\rangle \notin A$ or (iii) $\ulcorner\phi\urcorner = \ulcorner(\psi \wedge \chi)\urcorner$ and $\langle \ulcorner\psi\urcorner, s\rangle \in A$ and $\langle \ulcorner\chi\urcorner, s\rangle \in A$ or (iv) $\ulcorner\phi\urcorner = \ulcorner\exists x\psi\urcorner$ and there exists an assignment s' that is an x-variant of s such that $\langle \ulcorner\psi\urcorner, s'\rangle \in A$. See Murawski [1999, §2.6] for a similar approach and further details.

[10] A direct application of this method for a metatheoretic definition of satisfaction for second-order languages is given in Shapiro [1991, 135–136].

[11] In the second-order case, the metalanguage has to allow for quantification over arbitrary sequences of elements of *Rel*, or over arbitrary sets of relations on D.

[12] For details in the case of arithmetical languages, see Hájek and Pudlák [1993].

For this purpose, let $\mathcal{L}_\omega$ be a language in simple type theory. It contains denumerable sets V_τ of variables for each type τ: $x_1^\tau, x_2^\tau, x_3^\tau, \ldots$. It also contains, as non-logical vocabulary, a denumerable set of individual constants $a_1, a_2, a_3, \ldots$ and, again for each type τ (and arity n), denumerable sets of predicates $R_1^\tau, R_2^\tau, R_3^\tau, \ldots$ and functions $f_1^\tau, f_2^\tau, f_3^\tau, \ldots$.[13] Here the *type* of an expression is defined inductively: (1) i is the type of individual expressions; (2) the type of an n-ary relation with arguments of types $\tau_1, \ldots, \tau_n$ is $\langle \tau_1 \ldots \tau_n \rangle$; (3) the type of an n-ary function with arguments of types $\tau_1, \ldots, \tau_n$ and values of type τ is $\langle \tau_1 \ldots \tau_n \colon \tau \rangle$. The corresponding *order* of an expression is defined as follows: (1) expressions of type i have order 0; (2) expression of type $\langle \tau_1 \ldots \tau_n \rangle$ have order 1+ the maximal order of expressions of type $\tau_1, \ldots, \tau_n$; (3) expressions of type $\langle \tau_1 \ldots \tau_n \colon \tau \rangle$ have order 1+ the maximal order of expressions of type $\tau_1, \ldots, \tau_n, \tau$.[14]

Turning to the semantics, let $\mathcal{M}^\omega = (\{D_\tau\}_\tau, I)$ be a standard (full) interpretation of $\mathcal{L}_\omega$. Here $\{D_\tau\}_\tau$ is a "frame," i.e., a set of domains for expressions of each type τ. I is an interpretation function that assigns to each constant expression of type τ an element from D_τ. (I assigns elements from D_i to the individual constants and corresponding subsets of D_τ to τ-typed predicates and functions.) Assignment functions $s_\tau \colon V_\tau \to D_\tau$ are specified accordingly.[15]

Finally, let $\mathcal{L}_n$ be the n-order sub-language of $\mathcal{L}_\omega$, where formulas can only contain constant expressions and free variables of orders $m \leq n$, as well as bound variables of order $m < n$.[16] Along such lines, we get a cumulative hierarchy of sub-languages $\mathcal{L}_\alpha$ contained in $\mathcal{L}_\omega$, for $\alpha = 1, 2, \ldots, \omega$. The relevant models $\mathcal{M}^n$ for $\mathcal{L}_n$ are substructures of $\mathcal{M}^\omega$ whose frames are restricted to domains of type τ with orders $l < n$. We can now extend $\mathcal{L}_\omega$ by the introduction of a series of satisfaction predicates $Sat^1, Sat^2, \ldots$, where Sat^{n+1} expresses the satisfaction relation between formulas of $\mathcal{L}_n$, models $\mathcal{M}^n$, and sequences of elements of the corresponding domain. They can be defined explicitly in analogy to the metatheoretic approach above. The resulting satisfaction predicates are typed and, as such, belong to the hierarchy of orders of the language.[17] Crucially, for every sub-language $\mathcal{L}_n$ a

[13]The reason for including all this non-logical vocabulary is to stay as close as possible to the historical type theories we will consider later. (Typed predicates and functions play an important role in modeling the semantics of axiomatic theories in them.)

[14]The formation rules can then be formulated accordingly. In particular, for an n-ary relation R of type $\langle \tau_1 \ldots \tau_n \rangle$ and terms $t_1, \ldots, t_n$ of type $\tau_1, \ldots \tau_n$, $R(t_1, \ldots, t_n)$ is an atomic formula. For a detailed presentation of type theory, see Andrews [2002, §50].

[15]See Andrews [2002, 185–186] for details. We consider only full models here.

[16]Thus, formulas in the second-order sub-language $\mathcal{L}_2$ considered earlier can contains as free variables individual variables, first-order relation variables, and second-order relation variables; but only individual variables and first-order relation variables can be bound.

[17]For a recent development of this approach, in a generalized setting, see Linnebo and Rayo [2012]. The basic idea was already discussed informally by Gödel in the early 1930s.

suitable satisfaction predicate Sat^{n+1} can be defined in $\mathcal{L}_{n+1}$. (This result depends on the fact that, instead of working with all variables of a given type, one can work with arbitrarily long finite sequences and that finite sequences $\langle x_1^{\tau_1}, \ldots, x_k^{\tau_k} \rangle$ can be coded in a single object x^τ whose order is that of the highest typed element in the sequence.) Note also that, for each $\mathcal{L}_n$, the definition of Sat^{n+1} can only be given in the essentially richer language $\mathcal{L}_{n+1}$.[18]

Such a stratified theory of semantic predicates is not metatheoretic in the sense of specifying distinct "Tarskian metalanguages" that contain a translation of the expressions of the object language in question. Instead, here the "metalanguages" for sub-languages of given orders are embedded in the higher orders of $\mathcal{L}_\omega$.[19] Along such lines one can argue that type theory can represent its own semantics, after all. But again, it is clear that the outlined method has inherent limits. In particular, it does not allow for the specification of a *single* satisfaction predicate for all formulas of $\mathcal{L}_\omega$. What we get, instead, is a hierarchy of typed partial satisfaction predicates for its sub-languages, parallel to the metatheoretic "Tarski-hierarchy".[20]

§3. Early model theory, 1930–1940. We turn now to the historical part of this paper, i.e., to several contributions to formal semantics made in the 'long' 1930s. Our main objective will be to establish to what extent the modern definitions of *satisfaction*, *truth*, and *general validity* were prefigured then. Specifically, we will compare six different attempts to define semantic notions from this period, presented in works by Carnap, Tarski, Gödel, and Hilbert.

As already mentioned, in the existing literature these approaches have been taken to be incompatible due to the different logical traditions—the type-theoretic tradition and the first-order tradition—from which they come.

Compare Gödel's correspondence with Carnap in 1932 (Gödel [2003]), its discussion in Awodey and Carus [2007], and Section 4.2 below.

[18]If we indicate orders by superscripts and let $n = m + 1$, the explicit definitions will look as follows: $Sat^{n+1}(\varphi, \mathcal{M}^n, s^m) \Leftrightarrow_{df} \forall X^n\, (\Phi(\mathcal{M}^n, X^n) \rightarrow X^n(\phi, s^m))$. The definition of Φ will have this form: $\Phi(\mathcal{M}^n, X^n) \Leftrightarrow_{df}$ for all formulas ϕ of $\mathcal{L}_n$ and all assignments s^m for $\mathcal{M}^n$, $X^n(\phi, s^m)$ iff either (i) ϕ is an atomic formula of the form $R(t_1, \ldots, t_n)$ and $\langle \|t_1\|_s^{\mathcal{M}^n}, \ldots, \|t_n\|_s^{\mathcal{M}^n} \rangle \in I(R)$ or (ii) ϕ is a negated formula of the form $\neg\psi$ and not $X^n(\psi, s^m)$ or (iii) ϕ is a conjunctive formula of the form $\psi \wedge \chi$ and $X^n(\psi, s^m)$ and $X^n(\chi, s^m)$ or (iv) ϕ is an existentially quantified formula of the form $\exists v \psi$ and there exists an assignment t^m for $\mathcal{M}^n$ that is an v-variant of s^m such that $X^n(\psi, t^m)$. See Linnebo and Rayo [2012, 32–34] for details, especially concerning the orders.

[19]Compare Church on this point: "L_{n+1} is a semantical meta-language of L_n, containing the semantical predicates that are applicable to L_n and containing also L_n itself plus additional r-types of free variables and additional r-types of bound variables that are not present in L_n." Church [1976, 104].

[20]Once more, Church emphasized the parallel: "Moreover it is quite indifferent whether we speak of a single language L and a hierarchy of orders of variables and predicates within it or whether we speak of an infinite hierarchy of languages $L_1, L_2, L_3, \ldots$ as it is evident that the distinction is only terminological." Church [1976, 104].

More specifically, it has been argued that the corresponding approaches to explicate notions like truth or validity directly reflect two incommensurable conceptions of logic—the "universalist" and the "model-theoretic" conception—that underlie these two traditions.[21] It is also sometimes held that, while the early proponents of the model-theoretic tradition (like Gödel and Hilbert) succeeded in defining more or less the modern semantic notions, "universalists" (like Carnap and, to some extent, the early Tarski) were prevented from doing so by their own background assumptions.

Both claims need to be qualified or even revised more substantively, as we will argue. On the one hand, it will turn out that there are several (sometimes surprising) points of contact between the six historical positions we will consider—similarities that *connect* the two logical traditions. On the other hand, closer inspection of the relevant definitions will lead to a number of aspects in which they all diverge from their modern counterparts—again *across* the two logical camps. A general reflection on both the similarities and the differences with modern notions will conclude this Section.

3.1. Carnap on *general axiomatics*. Rudolf Carnap's earliest contributions to formal or general axiomatics can be found in three places: an originally unpublished typescript, *Untersuchungen zur allgemeinen Axiomatik* (Carnap [2000]) that was essentially finished in 1928; two related journal articles, "Bericht über Untersuchungen zur allgemeinen Axiomatik" (Carnap [1930]) and "Über Extremalaxiome" (Carnap and Bachmann [1936]); and Part 2 of his textbook, *Abriss der Logistik*, where axiom systems of set theory, topology, projective geometry, Peano arithmetic, etc. are discussed (Carnap [1929, 70–79]).[22]

The background logic Carnap uses for formalizing mathematical theories, and for articulating several metalogical notions like *isomorphism*, *categoricity*, and *semantic completeness*, is a simple type theory similar to the interpreted language $\mathcal{L}_\omega$ outlined in Section 2.2. This is called the "contentual basic system" (*inhaltliche Grunddisziplin*) by him. Axioms are then treated as n-ary propositional functions, while the mathematical primitives of a theory are represented by free variables. Hence axiom systems have the following logical form: $\Phi(X_1, \ldots, X_n)$, where the "primitive terms" $X_1, \ldots, X_n$ are usually relation variables of the relevant type; or in notation closer to Carnap's own: fX, where X stands for the tuple $X_1, \ldots, X_n$.

On that basis, Carnap formulates an initial definition of "satisfaction" or "truth" for an axiom system relative to a "model". This very early approach is developed in most detail in *Untersuchungen* (Carnap [2000, 95]), while the

[21]See again, e.g., Goldfarb [1979] and Hintikka [1988].

[22]For further background and discussions of Carnap's work on general axiomatics, see Awodey and Carus [2001], Awodey and Carus [2009], Bonk and Mosterin [2000], Goldfarb [2005], Reck [2004], Reck [2007], Schiemer [2012b], and Schiemer [2013].

published report on his "metalogical" results from 1930 contains a shortened version. The latter starts as follows:

> If fR is satisfied by the constant R_1, where R_1 is an abbreviation of a system of relations $P_1, Q_1, \ldots$; then R_1 is called a "model" of f. A model is a system of concepts of the basic system, generally a system of numbers (number classes, relations and so forth). (Carnap [1930, 303].)

Here "models" are not yet understood in the modern sense, but as sequences, or n-tuples, of (logical) relation constants of the interpreted background logic.[23] In *Untersuchungen*, Carnap is explicit about that aspect:

> [W]e speak of "models" of an axiom system and thereby mean logical constants, i.e., "*systems of concepts of the basic system*" (these are mostly systems of numbers). (Carnap [2000, 94].)

Given this understanding of models, the "satisfaction" of a theory by a model is treated in terms of the possible substitution of a constant $\mathcal{R}$, representing a model $\mathcal{M} = \langle R_1, \ldots, R_n \rangle$, for the "model variable" R of a theory.[24]

Carnap's basic semantic notion, "satisfaction", is treated informally at this point, i.e., no further specification of it is given either in *Untersuchungen* or Carnap [1930]. In particular, he does not formulate any conditions (apart from the "type adequacy" of models) that would resemble now standard recursive clauses for satisfaction or truth in a model. Nor does he distinguish satisfaction from truth (in the model-theoretic senses).[25] These aspects make it difficult to compare Carnap's earliest definitions of semantic notions, from 1928 and 1930, with the modern notions. That changes with his adoption of a metalogical stance in *Logische Syntax der Sprache* (Carnap [1934]), as we will see. (The issue how Carnap's initial work in logic is related to modern metalogic will be discussed further in Section 4.1.)

3.2. Gödel's dissertation. We now want to compare Carnap's very early semantic approach with the one presented by Gödel in his doctoral dissertation, "Über die Vollständigkeit des Logikkalküls" (Gödel [1929]).[26] The

[23]See Schiemer [2013] for a detailed discussion of Carnap's early model theory.

[24]The technique of treating satisfaction or truth *substitutionally* was typical at the time. As we will see below, it is also present in Gödel and Tarski's work from the same period. See Schiemer [2012b] for a further, more detailed discussion of its implications. Note also that there is a (partial) parallel to the use of added metatheoretic constants in defining satisfaction and truth that is standard in current model theory; compare footnote 6.

[25]There are indications that "being true" in type theory is understood by Carnap as equivalent to "being tautological" in *Untersuchungen*; see Bonk and Mosterin [2000, 38]. For more on the latter, including its Wittgensteinian roots, see Awodey and Carus [2009]. We will return to the further development of Carnap's approach in Section 3.4.

[26]As first pointed out in Awodey and Carus [2001], there had been a close exchange between Carnap and Gödel on logic in the late 1920s. In particular, Carnap gave a copy of his *Untersuchungen* manuscript to Gödel for comments. See also Goldfarb [2005], Reck [2007], Awodey and Carus [2009], and Reck [2013] for historical details.

direct background for Gödel is Hilbert & Ackermann's *Grundzüge der theoretischen Logik* from 1928, and specifically, the discussion of the "engere Funktionskalkül" and its possible completeness in it Hilbert and Ackermann [1928, 68]. Put in modern terms, the logic investigated by them is the restriction of type theory to first-order logic. Hilbert and Ackermann also introduce semantic notions such as "satisfiability" and "general validity", but without giving precise definitions.[27] In his dissertation, Gödel attempts to fills that gap. He provides the following definition of the notion of "satisfaction" for a first-order formula:

> Let A be any logical expression that contains the functional variables $F_1, F_2, \ldots, F_k$, the free individual variables $x_1, x_2, \ldots, x_l$, the propositional variables $X_1, X_2, \ldots, X_m$, and, otherwise, only bound variables. Let S be a systems of functions, $f_1, f_2, \ldots, f_k$ (all defined in the same universal domain), and of individuals (belonging to the same domain), $a_1, a_2, \ldots, a_l$, as well as propositional constants, $A_1, A_2, \ldots, A_m$. We say that this system, namely $(f_1, f_2, \ldots, f_k; a_1, a_2, \ldots, a_l; A_1, A_2, \ldots, A_m)$, *satisfies* the logical expression if it yields a proposition that is true (in the domain in question) when it is substituted in the expression. (Gödel [1929, 67–69].)

He adds:

> From this we see at once what we must understand by *satisfiable in a certain domain*, by *satisfiable* alone (there is a domain in which the expression is satisfiable), by *valid in a certain domain* (the negation is not satisfiable) and by *valid* alone. (Ibid.)

Several relevant differences to Carnap's approach can be noted immediately. First, "satisfaction in a system" is defined for first-order languages by Gödel, not for type theory. Second, whereas for Carnap the basic logical system, STT, is fully interpreted, the first-order logic specified by Gödel seems to be uninterpreted or formal in the modern sense. Interpretations, i.e., "systems" S, of the language can then be presented in an informal metalanguage. Third, Gödel, unlike Carnap, refers to an individual "domain" of the system to which the elements assigned to the individual variables belong and on which the respective relations and functions are defined. (This difference will be discussed further in Section 4.2.) Fourth, Gödel—like Hilbert & Ackermann before him, but going further than them—distinguishes explicitly between "satisfaction", "validity in a domain", and "general validity".[28]

At first glance Gödel's notions seem quite similar to the modern notions of satisfiability, truth in a model, and logical truth (Definitions 1–3 above).

[27] See, in particular, Hilbert and Ackermann [1928, §11].

[28] In the published version of his dissertation, Gödel specifies general validity as "'valid in every domain of individuals', which, according to well-known theorems, means the same as 'valid in the denumerable domain of individuals'." Gödel [1930, 103].

Yet on closer inspection several differences to the modern approach—and affinities to Carnap's—can be detected. Like Carnap, Gödel uses the word "satisfaction" for what we would call "truth in a model". As in Carnap's case, satisfaction in Gödel's sense is specified not relative to an interpretation and corresponding assignment functions (in the now standard sense), but in terms of "substitution" for variables.[29] Gödel then defines the satisfaction of a formula (or propositional function) in a system in terms of the "truth" of a resulting sentence (or proposition) and, again like for Carnap, the latter is left informal. Finally, "models" seem to be specified once more for formulas of a pure logical language, one without non-logical terminology.[30]

Interestingly, in the last section of his dissertation Gödel discusses briefly how his definition of satisfaction can be modified in order to yield a corresponding notion for axiomatic theories expressed in "applied languages" too, i.e., languages with non-logical "applied expressions."[31] Instead of specifying an interpretation function (an analogue to the modern I) that assigns elements of the relevant domain to the non-logical vocabulary of such a language, he proposes a technique that makes his approach strikingly similar to Carnap's (and, as we will see, to Tarski's). Basically, satisfaction of an applied system is specified in terms of translating the corresponding theory into a pure logical theory. More specifically, the semantics of applied axiom systems is expressed in terms of the semantics of pure systems by the "variabilization" of the mathematical terminology. In Gödel's own words:

> In order to prove that every applied first-order axiom system (whether finite or infinite) either has a model or is inconsistent, it suffices to consider the system of logical expressions that results from the applied axiom system when we replace the names by free variables and the functional constants by functional variables [...]. Now, if this system of logical expressions is satisfiable, then so is the axiom system. (Gödel [1929, 101].)

For further clarification, we can reconstruct this approach in modern terminology. Satisfaction is defined here for theories expressed as formulas of the following form: $\Phi(F_1, \dots, F_n, x_1, \dots, x_m, X_1, \dots, X_i)$. Like in Carnap's

[29]Note, once more, that one can see a partial parallel between this early substitutional approach and the use of added metatheoretic constants in modern model theory. (The relevant similarities and differences await further historical investigation.)

[30]Then again, for Gödel, unlike for Carnap, the formulas considered are first-order; they can contain free individual and function variables, but only bound individual variables.

[31]Applied languages are understood thus: "In order to deal with axiomatic questions, we require some additional notational conventions [...]. To the primitive signs we adjoin a denumerable sequence of *individual constants* (*names*), $a, b, c, \dots,$ and one of *functional constants* (*signs for notions*), $f, g, h, \dots$." Gödel [1929, 69] Gödel defines a mathematical theory, or an "*applied axiom system*", as a "(finite or infinite) system of applied expressions." (Gödel [1929, 69].) Hilbert's axiomatization of Euclidian geometry without the (second-order) axiom of completeness is mentioned as an example.

case, the primitive terms are represented by free variables.[32] Also putting aside Gödel's "substitutional" talk, a model or "system" for the theory can be understood as an ordered sequence $S = \langle f_1, \ldots, f_n, a_1, \ldots, a_m, R_1, \ldots, R_i \rangle$ that is assigned to the free variables in Φ. For Gödel, there is a common domain of individuals, say D, such that: (i) for each a_i: $a_i \in D$; (ii) for each relation R_i (of arity n): $R_i \subseteq D^n$, and (iii) for each (n-ary) function f_i: $D^n \to D$. Consequently, we can say that the system S *satisfies* the formula Φ if and only if $\Phi[S]$ is true in the given domain of individuals D.

Gödel's derivative notions can be reconstructed as follows: A theory T is *satisfiable* in domain D iff there exists an assignment S such that $\Phi[S]$ is true in it; or if one understands D as a model (for a language without non-logical vocabulary), iff $D \models \Phi[S]$. T is *satisfiable* iff there exists a domain D and an assignment S on it such that $D \models \Phi[S]$. T is *valid in* D iff for every possible assignment S on it, we have $D \models \Phi[S]$. Finally, T is *valid* iff for every domain D and for all assignments S on it, we have $D \models \Phi[S]$. Thus understood, Gödel's semantic notions are certainly more precise than those in Carnap's *Untersuchungen*. They are also closer to the current ones (as outlined in Section 2); but they are not identical with them. Most centrally, modern model theory is based on an explicit specification of the semantic ties between the schematically understood non-logical vocabulary of a language and a specific model, something not yet present in Gödel [1929].

3.3. Tarski on the concept of truth. In his long essay, "Der Wahrheitsbegriff in den formalisierten Sprachen" (Tarski [1935]), Tarski gives the first explicit metatheoretic definition of "truth" for a formal language.[33] However, he does not yet present it as a method for arbitrary formal languages; he only illustrates his approach for sentences of a specific "language of the calculus of classes" (LCC). The definition of truth for LCC is given in a richer metasystem, the "metacalculus of classes." (See Section 4.1 for Tarski's conception of metalanguage.) It is also indirect insofar as it is based on a prior definition of the "satisfaction" of sentential functions relative to "infinite sequences of objects". In that respect, he clearly goes beyond Carnap's and Gödel's earlier approaches.

Yet several points of contact with Carnap, in particular, should be noted as well. First, Tarski's notion of sentential function corresponds roughly to Carnap's notion of propositional function. In modern terminology, both can be represented by open formulas. Second, while Tarski's logical language LCC is not a type-theoretic language in the above sense, in an earlier, closely related essay on semantic definability, "Sur les ensembles définissables de nombres réel, I" (Tarski [1931]), he presented essentially the same definition

[32] Except, again, that Carnap considers only relation variables in *Untersuchungen*.

[33] The details and philosophical implications of Tarski's approach to truth have been intensively discussed in the literature. See Mancosu, Zach, and Badesa [2009] for further references.

of satisfaction for formulas of a "system of mathematical logic" derived from *Principia Mathematica*, namely a version of simple type theory that closely resembles Carnap's STT.[34] Third, and most importantly for our purposes, both object languages—STT in Tarski [1931] and LCC in Tarski [1935]—are explicitly described as fully interpreted languages, i.e., as meaningful formalisms, not as formal languages in the modern sense.[35]

Concerning the third point, in his 1931 paper Tarski points out that (i) the constant expressions of the type-theoretic language have a fixed interpretation and (ii) the variables of each type range over a domain of objects of a specific "order." More specifically, the domain of order zero consists of the real numbers, that of order one of "classes or sets" of real numbers, the next of classes of classes of real numbers, and so on (Tarski [1931, 113–114]). If formalized, this semantic conception of the universe of types would look very much like a model $\mathcal{M}^{\omega}$ for the interpreted type-theoretic language $\mathcal{L}_{\omega}$ described above. Given such a semantic understanding of STT, Tarski's "(infinite) sequences of objects" are then ordered sequences of objects of a specified type that are assigned to the variables of a formula.[36]

In his 1935 essay, using both sentential functions and (infinite) sequences, Tarski gives a "general definition of satisfaction" that essentially conforms to our inductive definition in Section 2.1 (Tarski [1935, 193]). In a footnote, he also explains how the inductive definition can be transformed into an explicit definition of the form: $Sat(x, s) \Leftrightarrow_{df} \forall X\, (\Phi(X) \rightarrow X(x, s))$, where Φ is closely related to the corresponding condition in our explicit definition of satisfaction in Section 2.2. Given his satisfaction predicate for formulas, *truth* for sentences is then defined thus: $T(x) \leftrightarrow_{df} \forall s\, (Sat(x, s))$ Tarski [1935, 195]. It should be emphasized, for our purposes, that satisfaction and truth are here specified relative to the fixed interpretation of the languages STT and LCC. In other words, what is defined is satisfaction for higher-order formulas by sequences of objects *in a given interpretation* of the object language in question; and accordingly, truth amounts to truth in a fixed,

[34]In the 1931 essay, Tarski works with finite sequences of objects, not with infinite ones, to define satisfaction; see Tarski [1931, 117]. In the 1935 essay, he explains both notions as follows: "The concept of *sequence* will play a great part in the sequel. An *infinite sequence* is a one-many relation whose counter-domain is the class of all natural numbers excluding zero. In the same way, the term '*finite sequence of n terms*' denotes every one-many relation whose counter-domain consists of all natural numbers k such that $1 \leq k \leq n$ (where n is any natural number distinct from 0)." (Tarski [1935, 171].)

[35]Compare Tarski on the meaningfulness of LCC in his 1935 paper: "[W]e are not interested here in 'formal' languages and sciences in one special sense of the word 'formal', namely sciences to the words and expressions of which no meaning is attached. For such sciences the problem here discussed has no relevance, is not even meaningful. We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider." (Tarski [1935, 166–167].)

[36]Compare also Tarski [1935, 191–193].

intended model.[37] Put differently again, Tarski's truth predicate conforms to the notion specified in Definition 2 above, namely $\mathcal{N} \models \varphi[s]$, for a *fixed* interpretation $\mathcal{N}$ of the object language.

Despite this centrality of *absolute* truth, the model-theoretic notion of *truth in a model* familiar from formal semantics today did not go unnoticed in Tarski's 1935 essay. In fact, in a small section dedicated to the "methodology of the deductive sciences (in particular, in the work of the Göttingen school grouped around Hilbert)" Tarski presents what he describes as a "special case" of the "absolute" notion of truth introduced above: "This is the concept of *correct* or *true sentence in an individual domain a*" (Tarski [1935, 199]). It is first characterized informally, as follows:

> By this is meant [...] every sentence which would be true in the usual sense if we restricted the expression of the individuals considered to a given class a, or—somewhat more precisely—if we agreed to interpret the terms 'individual', 'class of individuals', etc., as 'elements of the class a', 'subclass of a', etc., respectively. (Tarski [1935, 199].)

After that, a formal definition of the "relativized concept" of "satisfaction in an individual domain a" for LCC is presented. It modifies Tarski's original definition insofar as the sequences of elements assigned to the individual variables of a formula are restricted to a particular set a (Tarski [1935, 199–200]). Note here that the set a is not viewed as an independent domain by Tarski, but as a subset of the intended domain of individuals, say D_0, of LCC.[38] The reference to a particular set a as the new, restricted domain of individuals is thus gained by effectively restricting the range of individuals in the interpreted object language to a subset of D_0.

Based on such relativized notions of satisfaction and truth, Tarski also presents several semantic notions for general axiomatics that, as he remarks, have so far not been formally defined "in spite of the great importance of these terms for metamathematical investigations" (Tarski [1935, 200]). Definition 25 of his essay introduces the notion of a "correct statement in a:"

> x is a *correct* (*true*) *sentence in individual domain a* if and only if [...] every infinite sequence of subclasses of the class a satisfies the sentence x in the individual domain a (Tarski [1935, 200]).

[37] This is reflected in the fact that Tarski introduces a binary satisfaction predicate $Sat(x, s)$, not a ternary one, and a unary truth predicate $T(x)$, not a binary one.

[38] In Tarski's own words: "The modification depends on a suitable restriction of the domain of objects considered. Instead of operating with arbitrary individuals, classes of individuals, relations between individuals, and so on, we deal exclusively with the elements of a given class a of individuals, subclasses of this class, relations between elements of this class, and so on. It is obvious that in the special case when a is the class of all individuals, the new concepts coincide with the former ones [...]." (Tarski [1935, 239].)

Definition 27 then characterizes "correctness in every domain a", as follows:

> x is a *correct* (*true*) *sentence in every individual domain* [...] if and only if for every class a, x is a correct sentence in the individual domain a (Tarski [1935, 200]).

How exactly do these notions relate to their modern counterparts, as specified in Section 2.1? Tarski's "satisfaction in a" for different sets a does closely resemble the notion of satisfaction in a model $\mathcal{M}$ under assignment s, as specified in our Definition 1. But once more, the two are not identical.

The now standard notion is based on the idea of actually changing the interpretation of the object language in question. Thus, it is a relative concept in the precise sense that the truth of a formula (or theory expressed in the language) is specified relative to different interpretations of the logical language. If looked at closely, this is not what Tarski does in his 1935 essay. Basically, the semantic relativization underlying Tarski's approach concerns the class of possible assignments to variables. Recall that assignments corresponding to a model are today understood as functions from the set of variables to the model domain, i.e., $s\colon V_{\mathcal{L}} \to D$. Along Tarski's lines, what is restricted is not the individual domain D_0 of the (fixed) interpretation $\mathcal{N}$ of LCC, but simply the codomain of possible assignment functions.

To be even clearer, Tarski's "relativized" semantic notions can be reconstructed as follows: Let $s|^a$ denote an assignment function on model $\mathcal{N}$ with codomain a. Then "satisfaction in a" is given by $\mathcal{N} \models \varphi[s|^a]$ for the fixed interpretation $\mathcal{N}$ of the language and variable sets a. Thus, "correctness in a" expresses not the modern notion of truth in $\mathcal{M}$, as specified in Definition 2, but rather: $\mathcal{N} \models \varphi$ under all possible (restricted) assignments $s|^a$. "Correctness in all a" can be reconstructed accordingly: for all $a \subseteq D_0$, $\mathcal{N} \models \varphi$ under all possible $s|^a$. Once more, this is close to but not identical with the modern notion of validity, which is based on the metatheoretic generalization over all models of a language. At least in Tarski [1935], Tarski did not conceive of these semantic notions in their now canonical sense yet.

3.4. Carnap on models and analyticity. An explicit and logically precise definition of *truth in a model* for axiomatic theories is also given in Carnap & Bachmann's "Über Extremalaxiome" (Carnap and Bachmann [1936]).[39] This paper presents a formal treatment of so-called "extremal axioms" and is a direct continuation of Carnap's work on general axiomatics in the *Untersuchungen* manuscript. The notions of "axiom system" and "model" (similarly for "consequence", "isomorphism", etc.) are treated in more or less the same way as in 1928—with one central difference: truth in a model is no longer defined in terms of an informal notion of truth in STT, but by using the metatheoretic notion of "analyticity".

[39] This subsection is based on work presented in Schiemer [2013].

Carnap's notion of "analyticity" for formulas of a type-theoretic language, as used here, emerged around 1932, in close exchange with Gödel and while Carnap was working on *Logische Syntax der Sprache* (Carnap [1934]), one of his most well-known books.[40] A formal version of it was first published in "Ein Gültigkeitskriterium für Sätze der klassischen Mathematik" (Carnap [1935]), then also in the expanded English translation of his book, *Logical Syntax of Language* (1937). The notion of "analytic in LII" presented in the latter has striking similarities to the definition of truth in Tarski [1935]. Despite Carnap's overall "syntactic" point of view at the time, it is genuinely semantic in nature. The conditions for the analyticity of formulas of LII are defined similarly to the modern inductive definition of satisfaction. Comparable to Tarski's method, the notion is specified relative to "valuations", i.e., assignments of objects to the free variables of the language.[41]

In *Logical Syntax*, Carnap's object language LII closely resembles our type-theoretic language $\mathcal{L}_\omega$. Like in Tarski's case, it has a fixed, intended model $\mathcal{N}$. This means that all constant expressions have a fixed extension of corresponding type and all variables range over a specified type domain. Carnap's intended interpretation for LII can again be understood in terms of a standard model $\mathcal{M}^\omega$ for $\mathcal{L}_\omega$. Unlike in Tarski [1931], the individual variables do not range over $\mathbb{R}$ but over a set (of "positions") isomorphic to the natural numbers, i.e., $D_0 \cong \mathbb{N}$ (see Carnap [1937, §38a]). A "valuation" is then conceived of as an assignment of a (properly typed) object from one of the type domains of $\mathcal{N}$ to a variable of LII.[42] Consequently, *analyticity* of a formula relative to a (set of) valuation(s) can be reconstructed in terms of the modern notion of satisfaction in a model relative to an assignment s:

$$\varphi \text{ is analytic in LII } \Leftrightarrow_{df} \mathcal{N} \models \varphi[s] \text{ for every } s.$$

The notion of "analytic in LII" is introduced by Carnap as the formal reconstruction of the previously informal notion of "logical truth". This is jarring from today's point of view, since we understand the latter in terms of truth under all interpretations (Definition 3). In contrast, for Carnap—as for Tarski in the 1930s—it meant truth in a single intended model under all valuations. (We will come back to this point below.)

[40] See Coffa [1991] and Awodey and Carus [2007] for historical details.

[41] For the details of Carnap's definition, see Creath [1990]. For a closer comparison with Tarski's approach, see Coffa [1991].

[42] To be exact, Carnap (unlike Tarski) still treats assignments to individual variables substitutionally in *Logical Syntax*, while higher-order variables are treated objectually in the standard way. (See Awodey and Carus [2007] for further details.) Carnap also does not employ finite or infinite sequences; he uses the terminology of "valuation" to mean assignments to single variables. However, in §34c of *Logical Syntax* (1937), while talking about valuations for variables of different types, he comes very close to Tarski's notion of ordered sequences: "A valuation of the type $t_1, t_2, \ldots, t_n$ is an ordered n-ad of valuations which belong to the types t_1 to t_n respectively." (Carnap [1937, 108].)

The definition of *model*, thus also of *truth in a model*, for axiomatic theories in Carnap and Bachmann [1936] is then based on his notion of analyticity in a type-theoretic language. The new definition is this:

> Let 'M_1' be an abbreviation for a sequence of constants of the language S. We say that M_1 is a model of the axiom system '$F_1(M)$' if the sentence '$F_1(M_1)$' is analytic in S [i.e., LII]. (Carnap and Bachmann [1936, 67].)

As in Carnap [1930], models are here conceived as sequences of relation constants, this time for the type-theoretic object language LII. A theory, expressed by the formula $F_1(M)$ with model variable M, is true in model M_1 if and only if the resulting sentence $F_1(M_1)$ is analytic in LII. Unlike earlier, Carnap's definition is now clearly metatheoretic. Yet what we get is again not the modern notion (as specified in Definition 4). In particular, truth in a model is defined for a theory formulated in a language with a fixed interpretation, while the modern model-theoretic definition presupposes the re-interpretability of the language in which the theory is formulated.

If we put aside once more the fact that model constants are "substituted" for the model variables (as we did earlier, also in Gödel's case), the following reconstruction of the difference seems plausible. For Carnap, models are tuples $\mathcal{M} = \langle R_1 \ldots R_n \rangle$, where each R_i is a relation defined on the individual domain D_0 of the language LII; i.e., models are effectively assignments to the variables of a formula. Given his treatment of axiomatic theories, truth in a model $\mathcal{M}$ for a sentence $\Phi(X_1, \ldots, X_n)$ of LII with "primitive signs" $X_1, \ldots, X_n$ is *not* the modern $\mathcal{M} \models \Phi$ but something like $\mathcal{N} \models \Phi(X_1, \ldots, X_n)[R_1 \ldots R_n]$. Since $\mathcal{N}$ is the fixed interpretation of LII, the only kind of variability possible in this context is in terms of the variation of assignments to the theory's "primitive variables" $X_1, \ldots, X_n$.[43]

Obviously Carnap's approach has strong similarities not only to Gödel's in his dissertation, but also to Tarski's, e.g., in Tarski [1935]. The latter becomes even more apparent if we compare Carnap and Bachmann [1936] with another famous paper of Tarski's published in the same year.

3.5. Tarski on models and logical consequence. Tarski's seminal paper, "Über den Begriff der logischen Folgerung" (Tarski [1936]), has long been considered the origin of the modern notions of *logical truth* and *logical consequence*.[44] The following definition of "model" for a sentence, or set of sentences, is presented in it:

> Let L be any class of sentences. We replace all extra-logical constants which occur in the sentences belonging to L by corresponding variables, like constants being replaced by like variables and

[43]See Schiemer [2012a] and Schiemer [2013] for more details.

[44]This view has recently been challenged in a number of studies. See Mancosu [2010a] for an overview of the debate. (We will return to this issue in Section 4.2.)

> unlike by unlike. In this way we obtain a class L' of sentential functions. An arbitrary sequence of objects which satisfies every sentential function of the class L' will be called a *model* or *realization of the class L of sentences*. (In just this sense one usually speaks of models of an axiom system of a deductive theory.) (Tarski [1936, 416–417].)[45]

Tarski is explicit that the notion of satisfaction used here is that of Tarski [1935]. Against that background, "logical consequence" is defined thus: Sentence φ follows logically from a set of sentences Γ if every model of Γ is also a model of φ (Tarski [1983, 417]). Note that the definition of a model used here is based on the same convention present in Gödel [1929], namely the "variabilization" of the non-logical constants and, thus, the translation of a mathematical language into a pure higher-order language.

There are also several striking similarities to Carnap & Bachmann's approach. Like them, Tarski works with axiomatic theories and their metatheory in a higher-order logical framework. Moreover, in both papers a definition of truth in a model is presented based on the prior specification of a formal notion of truth in a language with fixed interpretation—LII in Carnap [1934] and LCC in Tarski [1935]. Tarski does not mention Carnap and Bachmann [1936] in his paper. However, he refers to Carnap [1934] and Carnap [1935] to note the similarities between his approach and Carnap's. More specifically:

> Finally, it is not difficult to reconcile the proposed definition [of logical consequence] with that of Carnap. [...] Analogously, a class of sentences can be called analytical if every sequence of objects is a model of it. (Tarski [1983, 418].)

What is particularly interesting here is that Tarski, seemingly unaware of Carnap's updated definitions in his 1936 paper, attributes a conception of models to Carnap that he sees as equivalent to his own.[46] In particular, both logicians identify models with "sequences of objects" of (or relations on) the intended domain of individuals of the underlying language.[47]

A further similarity, not highlighted so far, concerns a second difference to the modern definition of models that is crucial for our purposes. As in Carnap and Bachmann [1936], there is no explicit sign indicating the domain of a model in Tarski's 1936 definition of models. In contrast to Tarski [1935], where the issue of a relativized notion of satisfaction and truth to a specific set is explicitly discussed, nothing is said here concerning

[45] A similar definition of a "model of an axiomatic theory" occurs in Tarski [1937].

[46] Besides Carnap [1934] and Carnap [1935], Tarski's definitely was acquainted with Carnap's work on axiomatics in *Untersuchungen* at this point. As first noted in Awodey and Carus [2001], he discusses it with Carnap in Vienna in 1930.

[47] See also Carnap's *Logical Syntax*, from 1937, §71e, where a model domain is defined as the "domain of the substitution-values of the [primitive] variable" (Carnap [1937, §71e]).

the specification of a domain. (We will explore this issue further in Section 4.2.)

3.6. Hilbert & Bernays' *Grundlagen*. Finally, we turn to the treatment of semantic notions in Hilbert & Bernays' monumental *Grundlagen der Mathematik*, Volumes I-II (published in 1934 and 1939, respectively), in particular to their definitions of "satisfiability" and "general validity".[48] Volume I introduces the notion of "satisfaction" in the context of formal axiomatics. Contrary to what one might expect, the treatment of axiom systems outlined in it is identical to Carnap and Tarski's: the primitive terms of a theory are expressed by free predicate variables, the axioms and the theory itself by propositional functions.

As an example, Hilbert and Bernays consider axioms for Euclidian geometry with two primitive predicate variables R (for *greater than*) and S (for *between*). The axiom system is symbolized as $\mathsf{U}(R, S)$, a model for it as a tuple of the form $\langle R_1, S_1 \rangle$. $\mathsf{U}(R, S)$ is called "satisfied" if there exists a pair of predicates $\langle R_1, S_1 \rangle$ such that $\mathsf{U}(R_1, S_1)$ is a "correct proposition" (Hilbert and Bernays [1934, 8]). The two authors go on:

> A formula is called generally valid if it represents a true sentence under *every* determination of the variable predicates; it is called satisfiable if it presents a true sentence under an *appropriate* determination of the variable predicates (Hilbert and Bernays [1934, 8, our translation, emphasis in the original]).[49]

Here the central notion is that of a "determination" of the free relation variables of the language. In Hilbert and Bernays' first attempt at a more detailed definition, which occurs in the context of a discussion of Gödel's completeness theorem for first-order logic in the same volume, "determination" (*Bestimmung*) is replaced by "substitution" (*Einsetzung*):

> A formula of the predicate calculus is called generally valid if it yields the value 'true' for every substitution of logical functions for the formula variables and, in case free individual variables occur, for every substitution of elements of the individual domain for the free individual variables. It is called satisfiable if it yields the value 'true' relative to an appropriate choice of substitutions (Hilbert and Bernays [1934, 128, our translation]).[50]

[48]They are closely related to corresponding definitions in Hilbert and Ackermann [1928]. For our purposes, it is illuminating to note the continuities and changes at this point.

[49]In the original German: "Eine Formel diese Art heisst allgemeingültig, wenn sie für *jede* Bestimmung der variablen Prädikate eine wahre Aussage darstellt, sie heisst erfüllbar, wenn sie bei *geeigneter* Bestimmung der variablen Prädikate eine wahre Aussage darstellt." (Hilbert and Bernays [1934, 8].)

[50]"Eine Formel des Prädikatenkalküls heisst allgemeingültig, wenn sie bei jeder Einsetzung von logischen Funktionen für die Formelvariablen und, falls freie Individuenvariablen vorkommen, von Dingen des Individuenbereichs für die freien Individuenvariablen den Wert

The process of substituting a "logical function" for an (n-ary) "formula variable" is specified further in terms of the assignment of a set of n-tuples of elements from the individual domain for which the formula is true (Hilbert and Bernays [1934, 126]).

In Volume II of *Grundlagen*, this approach is still further developed. Now the following definition of "satisfiability in domain J" for a formula $\mathfrak{F}$ of pure first-order logic is given:

> This means that, given a proper individual domain J, there exist such substitutions of elements of J for the free individual variables in $\mathfrak{F}$, of truth values for formula variables without arguments, and of logical functions with the same number of arguments for formula variables with arguments, relative to domain J, that the formula yields the value "true" [...] (Hilbert and Bernays [1939, 179]).[51]

The "general validity" of $\mathfrak{F}$ is then characterized accordingly:

> [A formula is generally valid if] under the replacement of the free individual variables by elements of the individual domain, of the formula variables without arguments by truth values, and of the formula variables with arguments by logical functions, relative to domain J, it always yields the value 'true' on the basis of interpreting the logical signs (Hilbert and Bernays [1939, 179, our translation]).[52]

Here the terminology of "replacement" (*Ersetzung*) is used instead of "substitution" (*Einsetzung*).

Clearly Hilbert & Bernays' treatment of "satisfiability" comes very close to the modern definition of satisfiability in a model relative to an assignment (Section 2.1, Definition 1). But as in the previous cases, there are differences as well. In particular, no models in the modern sense, including the use of interpretation functions for non-logical constants, are mentioned. The situation is more complicated with respect to the definition of "general validity". Note that validity as specified above is *not* the modern notion of truth in all models. Hilbert & Bernays' use of the phrase "under *every* valuation of the

'wahr' ergibt; sie heisst erfüllbar, wenn sie bei passender Wahl der Einsetzungen den Wert 'wahr' ergibt." (Hilbert and Bernays [1934, 128].)

[51] "Sie besagt, dass es für einen geeigneten Individuenbereich J solche Einsetzungen der freien Individuenvariablen in $\mathfrak{F}$ durch Individuen aus J, der Formelvariablen ohne Argument durch Wahrheitswerte und der Formelvariablen mit Argumenten durch logische Funktionen mit der gleichen Anzahl von Argumenten, bezogen auf den Bereich J, gibt, für welche die Formel den Wert 'wahr' erhält." (Hilbert and Bernays [1939, 179].)

[52] A formula is generally valid if "bei Ersetzung der freien Individuenvariablen durch Dinge aus dem Individuenbereich J, der Formelvariablen ohne Argument durch Wahrheitswerte und der Formenvariablen mit Argumenten durch logische Funktionen, bezogen auf den Bereich J, auf Grund der Deutung der logischen Zeichen stets den Wert 'wahr' erhält." (Hilbert and Bernays [1939, 179].)

variable predicates" in Volume I seems to indicate such an understanding. However, their definition in Volume II, as just quoted, makes evident that what is intended is validity relative to a single model or "individual domain J". Hence, what they define as "general validity" corresponds roughly to the modern notion of *truth in a model* (Definition 2), as opposed to *logical truth* (Definition 3). As such, it is equivalent to Carnap's notion of "analyticity in LII", as well as to Tarski's notion of "correctness in set a".

Having said that, it is important to acknowledge that Hilbert and Bernays do not generally confuse "general validity" with truth in all models. Just as Gödel distinguishes between validity in a single domain and validity in all possible domains in Gödel [1930], a similar distinction can be found in *Grundlagen*, even if only in passing. After a discussion of consistency proofs for axiomatic theories, the two authors explicitly distinguish between satisfiability in a denumerable domain, satisfiability in a non-denumerable domain, and "general validity in *every* domain of individuals." (Hilbert and Bernays [1934, 9])[53] In other words, Hilbert & Bernays are aware of the difference between the model-theoretic notions captured by our Definitions 2 and 3. The point to stress for our purposes is that, nonetheless, in the 1930s a sharp distinction between these two notions was generally not made.

3.7. Points of contact and similarities. In the previous subsections, six early attempts at defining genuinely semantic notions—like *satisfaction*, *truth in a model*, and *validity*—were presented. Some of them grew out of each other directly. Thus, Carnap & Bachmann's 1936 definition of truth in a model in terms of "analyticity in LII" is an attempt to explicate the corresponding informal notion in Carnap's earlier *Untersuchungen*. Likewise, Tarski refers to his definition of "satisfaction in a sequence" from Tarski [1935] when defining truth in a model and logical consequence in his 1936 paper. Finally, Gödel's notions of "satisfiability" and "validity" in his dissertation are explicitly influenced by Hilbert's prior work, in particular Hilbert and Ackermann [1928].

Beyond such connections, the accounts we discussed are usually assumed to constitute independent, partly even incompatible contributions to the field. Our close inspection has revealed, however, that there are further points of contact between them. In fact, the semantic notions introduced by Carnap, Tarski, Gödel, and Hilbert in the 1930s are similar in several striking respects, despite the different "logical frameworks"—type theory *versus* first-order logic—in which they are formulated. Moreover, these similarities distinguish them from their modern counterparts. To conclude this Section, we now summarize the five most important similarities. In doing so, we also probe some of them in more depth.

[53]A similar distinction already occurs in Hilbert and Ackermann [1928], also in passing: "Genauer müsste man statt Allgemeingültigkeit Allgemeingültigkeit für jeden Individuenbereich sagen." (Hilbert and Ackermann [1928, 95].)

1. A characteristic shared by the accounts is their focus on the formal semantics for axiomatic theories.

In all our six cases, the approach is motivated to a large degree as an attempt to make more precise semantic notions that had previously been used informally in the context of mathematical axiomatics.[54] This is clearly so for Carnap's project of "general axiomatics", in his *Untersuchungen* and subsequent writings. It is also the case in Tarski's work on the "methodology of the deductive sciences", for instance in the section on the relativized satisfaction concept in his 1935 paper. Moreover, recall that in Tarski [1936], after specifying his conception of formal models, he states that "in just this sense one usually speaks of models of an axiom system of a deductive theory" (Tarski [1983, 417]).[55] The same holds for Gödel's discussion of "applied languages" in the formalization of theories in Gödel [1929] and, as we just saw, for Hilbert & Bernays' notions of satisfiability in both volumes of *Grundlagen*.

2. All six accounts contain essentially the same technique for formalizing axiomatic theories.

The relevant convention, in all the works considered, is that of treating the primitive mathematical vocabulary of a theory in terms of variable expressions. Axiomatic theories are formalized as sentential functions, i.e., open formulas in the modern sense. The primitive terms of a theory are thus represented as (higher-order) variables and not, as is standard practice today, in terms of non-logical constants. Consequently, semantic notions like satisfaction or truth in a model are defined for pure logical languages, without non-logical terminology. In cases where genuinely mathematical or "applied languages" are considered, as in Gödel [1929] and Tarski [1936], the semantic evaluation of sentences is done indirectly, via the method of "variabilization", i.e., the translation into a pure language where free variables (of the right type) are substituted for the non-logical constants.

This point deserves further elaboration. In the literature—most explicitly and prominently in Hodges [1986] and Demopoulos [1994]—in has been argued that this convention of treating primitive terms as variables marks a clear difference to the modern model-theoretic understanding of languages. From this point of view, an innovation necessary, or at least highly conducive, for the emergence of model theory was the introduction of the modern notion of non-logical constants, or more precisely, the new semantic understanding of them *as schematic terms* that is still missing in the accounts presented in this paper.[56] Note here that several of the logical languages mentioned above

[54]Compare Awodey and Reck [2002] for a survey of contributions to the (meta-)theory of formal axiomatics in the late 19th and early 20th centuries.

[55]See Jané [2006] for more on the axiomatic background of Tarski's paper.

[56]The terminology of "schematic terms" in this context, while not widespread in mathematical logic, is fairly common in the philosophical literature; see, e.g., Quine [1986].

do contain constant vocabulary, i.e., individual constants (such as numerals), predicates, function symbols etc.; language LII in Carnap [1934] and Gödel discussion of "applied languages" are cases in point. What distinguishes them from "model-theoretic" languages is the specific semantic treatment of these constants. In "Frege–Peano languages", like LII, they have a fixed extension in the intended interpretation of the language. In modern model theory, in contrast, they have a fixed interpretation only relative to a given structure, in the now standard sense.

But noting the predominance of the method of "variabilization" in the 1930s, this issue deserves to be revisited. Why exactly should the mathematical primitives of a theory not be formalized in terms of free (higher-order) variables, as outlined above, instead of treating them as non-logical constants? For Hodges and Demopoulos, the crucial issue is the different *semantic roles* of variable and constant expressions. Namely, (i) non-logical constants are "expressions with a determinate meaning, but with no fixed reference," whereas (ii) variables are "expressions lacking any meaning while admitting a conventionally stipulated reference (within a given 'range')" (Demopoulos [1994, 214]). The crucial difference lies, thus, in their roles in the truth evaluation of sentences of the language (again relative to a model $\mathcal{M}$). As both authors stress, the truth of a sentence depends primarily on the semantic value of the constants (relative to $\mathcal{M}$), while the assignments play only a secondary role.[57] Thus, model theory depends on expressions that can be reinterpreted in different models, but have a fixed extension relative to a given model. Variables usually don't work this way.[58]

Now, let us reconsider how this issue plays itself out in the various semantic accounts developed in the 1930s. It is closely related to another shared characteristic of the approaches we considered.

3. In the 1930s, models are usually treated as assignments to the "primitive variables" of a theory.

As highlighted above, logicians in the 1930s still had a different understanding of models compared to the present one. Models were usually taken to be

[57]Demopoulos puts this point as follows: "In the course of evaluating the truth of a sentence in a structure, we may stipulate a reference for the variables relative to an arbitrary sequence of elements of the domain of the structure. But we do not regard the sequence an essential feature of the structure. Not so for the distinguished elements, relations, and operations which interpret the non-logical constants—they *are* an essential feature of the structure." (Demopoulos [1994, 215].)

[58]If one strips away the intensional talk of "meaning", the semantic difference boils down to this: Given a model $\mathcal{M}$, there exists (by definition) an interpretation function I that assigns to each constant an element of the domain, a relation on the domain, a function on the domain, etc. But given the same $\mathcal{M}$, there exist usually many different assignment functions s that assign to a given variable an element of D, a relation on D, or a function on D. Consequently, the extension of a constant is fixed relative to $\mathcal{M}$, while the extension of a variable can, and usually does, still vary.

sequences of objects, relations, and functions that are assigned to the primitive terms of a theory expressed as free variables. Especially in the context of formal axiomatics—and *instead* of using a domain and corresponding interpretation function—models here simply consist of assignments to such "primitive variables" of a theory. Evidently, these assignments are constitutive for the specification of models, similarly to how an interpretation function is constitutive for a structure in modern model theory.

But then, do the historical approaches surveyed by us really differ as strongly from modern model theory as suggested by Hodges and Demopoulos? Again, these two authors emphasize the distinction between two levels of semantic generality present in the modern account: constants have a fixed interpretation relative to a given model, while variables remain semantically flexible even after its specification. It is this distinction that is assumed to be missing before the advent of modern model theory. Yet, if one looks closely at how the primitive terms of axiomatic theories are evaluated by Carnap, the early Tarski, etc., it becomes evident that a parallel distinction between two levels of generally is intended by them. In particular, the "primitive variables"—as a specific class of variables used in the formalization of a theory—play essentially the same semantic role as "schematic" constants in the modern treatment. They are re-interpreted from model to model, but take on a fixed reference relative to a given interpretation of the theory. This is not the case for the "non-primitive" variables.[59]

Hence, the specific (higher-order) variables used to represent the mathematical vocabulary of a theory in our historical accounts do not function semantically like free variables in model theory. Instead, their semantics can be understood more properly in terms of the usual understanding of a particular kind of mathematical variables, namely, that of *parameters*. Basically, parameters are variable expressions of a language that are assigned a fixed value in a given context or a given example. From a model-theoretic perspective, their semantic role is similar to that of "schematically" understood non-logical constants: their extension is fixed relative to a given model (while different assignments to the other variables are still possible), yet it can vary from model to model.[60] But given this understanding of the primitive terms of a theory, the claim by Hodges and Demopoulos that there is a decisive shift in the move from our early semantic accounts to modern model theory is undermined. Our discussion suggests a more *continuous* evolution from the "Frege–Peano languages" to modern "model-theoretic languages."

[59]Note that with an assignment to the primitive variables, the other, more usual variables of the language are not interpreted yet. Models of theories, in the sense at issue, do not include assignments to the latter, except that they are constrained to the entities assigned to the primitive variables.

[60]When characterizing the semantics of non-logical constants in modern model theory, Demopoulos makes a closely related point: "Thus, as I am using the term, non-logical constants correspond to what are sometimes called parameters." (Demopoulos [1994, 213].)

Two further similarities between our historical accounts should be highlighted. The next concerns a detail in the treatment of models.

4. In the accounts considered by us, model domains are usually not specified explicitly.

Given that models are standardly treated in terms of assignments to "primitive variables" in the 1930s, there remains a significant difference to modern model theory. Namely, in those early accounts the domains underlying such assignments are often not indicated directly. This is evident in Carnap's work on formal models from the period. Note, for instance, that in Carnap [1930] models are specified as sequences of relation symbols $\mathcal{M} = \langle R_1, \dots, R_n \rangle$, where each predicate letter R_i denotes a n-ary relation, say $R_i^{\mathcal{M}}$. But the domain of individuals on which these relations are defined is nowhere made explicit. More formally, assuming that $R_i^{\mathcal{M}} \subseteq D^n$ for an n-ary relation, the domain D is left unspecified by him.[61]

Gödel is more explicit about the domain of a "system" in his dissertation from 1929. Recall that in his corresponding definition of satisfaction, as quoted above, he mentions the "domain in question" for a particular system. He also holds that all functions are to be "defined in the same universal domain" and that all individuals have to belong to that domain. But that is all we are told.[62] Likewise, model domains are not explicitly specified in Carnap's account in Carnap and Bachmann [1936], nor in Tarski's 1936 paper on logical consequence. Particularly with respect to Carnap's and Tarski's approaches, the question arises, therefore, how the model theory of axiomatic theories expressed in languages like STT or LII is related to the background semantics, i.e., the intended interpretation of the language itself. (We will return to this issue in Section 4.2.)

Finally, the following trait is shared by our six accounts as well:

5. The distinction between truth in a model and logical truth (i.e., truth in all models) is often not sharply drawn.

The predominant conception of models as assignments to "primitive variables", together with the missing explicit specification of model domains in the 1930s, did have a negative effect. Typical for the period is a certain amount of confusion, or at least ambiguity, concerning some metalogical notions precisely demarcated today. In particular, no sharp line is drawn yet—by several leading logicians—between the notion of truth in a model (in the sense of our Definition 2) and logical truth (Definition 3). In some cases, "general validity" is simply defined in terms of truth in a model.

[61] See Schiemer [2013] for a detailed discussion of Carnap's conception of model domains in the context of his early semantics.

[62] As in Carnap's account, Gödel does not explicitly specify the individual domain for a given system $(f_1, \dots, f_n; a_1, \dots, a_m; A_1, \dots, A_i)$.

To be more specific again, Tarski's notion of "correctness in all [sets] *a*" in Tarski [1935] is not identical with the modern notion of logical truth as specified in our Definition 3. In Carnap's case, analyticity in LII does not correspond to Definition 3 either, but to Definition 2, even though it was intended to capture the informal notion of logical truth. What is perhaps more surprising, given the usual historical assumptions, is that even in Hilbert & Bernays' *Grundlagen* the notion of "general validity" is not unequivocally defined in terms of truth in all models (or domains) but, at least at certain places, as validity relative to all assignments from a single domain.

§**4. Further comparison.** In this section, two issues will be pursued further that have been prominent in the recent literature on the history of modern logic, namely:

1. To what extent are the semantic notions from the 1930s metatheoretic in the modern sense?
2. Do the corresponding approaches to model-theoretic truth allow for domain variation?

Our discussions so far led to several revisionary results concerning both issues. But more can be said in connection with each.

4.1. Monolinguism vs. metatheory. As we noted, Carnap's earliest discussions of notions like isomorphism, categoricity, and semantic completeness, in *Untersuchungen* and related writings, are presented in a single, "universal" background theory (his so-called *Grunddiziplin*). While he does introduce the notion of satisfaction as a predicate for sentences of his type-theoretic language in those texts, it is treated informally and clearly not in a metatheoretic way. Moreover, Carnap's project of general axiomatics is often criticized exactly for the missing distinction between object- and metatheory. For instance, in Coffa [1991] it is characterized as a "monolinguistic approach", thereby stressing its Russelian background.[63] Coffa emphasizes the difference to Tarski's later "metalinguistic" approach, as presented in Tarski [1935]; and he concludes that Carnap, because of his monolinguism, is "incapable of providing an adequate framework for either syntactic or semantic metamathematical notions" (Coffa [1991, 279]).[64] One might add, charitably, that it is understandable why Carnap did not built his early model theory on such a systematic distinction. Prior to Gödel's and Tarski's limitative results from the early 1930s there was simply not much motivation for an explicit separation of object- and metalanguage.

[63] In Coffa's words: "Carnap's book was thus inspired by the somewhat epicyclic aim of showing that everything of value in metamathematics can (or should) be expressed within the monolinguistic framework of *Principia Mathematica*." (Coffa [1991, 274].)

[64] Accounts of Carnap's monolinguistic stance that are less critical than Coffa's and emphasize Carnap's genuine contributions to metalogic are Awodey and Carus [2001], Awodey and Reck [2002], Reck [2007], and Reck [2013].

Our discussion so far suggests a further defense of Carnap and a different perspective overall. While his early approach is clearly not metatheoretic in the later Tarskian sense, it is not "monolinguistic" in Coffa's sense either. In fact, an object-/metalevel distinction is simulated in his background theory. Carnap's *Grunddisziplin* is really separated into two distinct parts with different functional roles: a *pure* part and a *contentual* part. The simulation of an object-/metalevel distinction within a single type theory is gained by the convention to formalize axiomatic theories in the pure sublanguage, where the basic axiomatic concepts are represented by higher-order variables. The "metatheory" is specified in the extended background language, which includes logically defined, "absolute" constant expressions from relation theory, set theory, and arithmetic (Carnap [2000, 60]).[65] This "duplication" of terms shows that Carnap was aware of the need to distinguish language levels—particularly for treating axiomatic theories semantically—even within a universal language like STT. His "monolinguistic" approach should thus be viewed as a *predecessor* of Tarski's introduction of semantic metalanguages, not as radically different from it.[66] It is also less limited than often assumed.

Carnap's contributions to the development of modern logic are still often neglected, since they are assumed to be idiosyncratic and clearly inadequate. This is obviously not the case for Tarski, Gödel, and Hilbert. But were their approaches really that far from Carnap's in the 1930s, especially with respect to the issue of "metatheory"? Gödel's and Hilbert & Bernays' accounts of "satisfiability" and "general validity" for pure first-order logic are very similar to each other, as we saw. The definitions presented in their works are also clearly metatheoretic from a modern perspective. Thus, in Volume II of *Grundlagen* the truth conditions for formulas of the pure predicate calculus (*reine Prädikatenkalkül*) are explicitly specified in a (nonfinitary) "set-theoretical predicate theory", i.e., a first-order set theory like ZF (Hilbert and Bernays [1939, 189]). Moreover, the two authors articulate a fully formalized definition of truth for Peano arithmetic based on the Gödelization of terms and formulas, one that is comparable to the definition given at the end of Section 2.1 above (Hilbert and Bernays [1939, 330–340]).

In Gödel's case it would be misleading, however, to consider only his treatment of first-order logic when analyzing his contributions to semantics from the 1930s. This becomes even clearer if we go beyond the writings already mentioned so far. For example, the logic discussed in his famous

[65] As Carnap writes: "Every treatment and investigation of an axiom system therefore presupposes a logic, specifically a contentual logic, i.e., a system of sentences that are not merely combinations of signs but that have a particular meaning." (Carnap [2000, 60].)

[66] See also Bonk and Mosterin: "Carnap's Reflexion über die Notwendigkeit einer Grunddiziplin in den *Untersuchungen* unterscheidet zwischen Mathematik und dem Diskurs über Mathematik. Das Verhältnis von Grunddiziplin zu Axiomensystem ist hier analog zur Beziehung von Metasprache und Objektsprache [. . .]." (Bonk and Mosterin [2000, 35].)

incompleteness paper, "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I" (Gödel [1931]), is a version of simple type theory very similar to Carnap's and to that outlined in Section 2. Now, there is a famous footnote (48) in that paper where the hierarchy of types and its relation to questions of truth and decidability is addressed:

> As will be shown in Part II of this paper, the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite [. . .], while in any formal system at most denumerably many of them are available. For it can be shown that the undecidable propositions constructed here become decidable whenever appropriate higher types are added. (Gödel [1931, 181].)[67]

The second part of Gödel [1931], as announced in this passage, was never realized. But this and related remarks—on the decidability of sentences and the definability of truth in higher "type extensions"—indicate that what Gödel envisaged was a *stratified* account of truth close to that outlined in Section 2.2., i.e., a hierarchy of truth predicates in a single type theory.[68]

Tarski [1935] is generally viewed as the *locus classicus* of a metatheoretic account of formal semantics. Tarski provides the first detailed specification of the necessary constituents of a metatheory in which a truth predicate can be defined. The metalanguage has to contain (i) a translation of the expressions of the object language, (ii) a description of its syntax, and (iii) sufficient mathematical resources to express the satisfaction clauses specified above (Tarski [1935, 171]). In the last two sections of the essay, as well as in a Postscript added later, the necessary resources of a "Tarski metalanguage" are specified for formal languages more generally. A central concept introduced at this point is that of the "order of an expression", corresponding to its "semantic category" or type (Tarski [1935, 218]).[69]

Given this specification of orders, §4 of Tarski's long essay contains an interesting specification of the satisfaction relation for object languages of order n, where $2 \leq n \leq \omega$. He shows that a "correct definition of satisfaction" for object languages with "more complicated logical structure" than those of first order has to reflect the different orders of formulas and the different semantic domains of the language from which sequences of objects

[67]See Feferman [2008] for a discussion of the so-called "Gödel's doctrine" and, more generally, his "caution" concerning the use of set-theoretic truth in this context.

[68]In addition, Gödel's subsequent work on the constructible universe L is directly motivated by his reception of Russell's ramified type theory and by the idea of a stratified account of definability. See Kanamori [2007] for a detailed reconstruction of the transition in Gödel's work from truth in type theory to higher axiomatic set theory.

[69]In his specification of orders of expressions and languages, Tarski refers to Carnap's *Abriss* and the discussion of simple type theory in it; see §13, entitled "The hierarchy of types" Carnap [1929, 218]. Carnap's and Tarski's accounts essentially conform to the discussion of types and languages of orders $= 1, 2, \ldots, \omega$ given in Section 2.2.

can be drawn (Tarski [1935, 224–225]).[70] He thus proposes a modified definition of satisfaction for formulas with mixed type structures in terms of the "method of many-rowed sequences" (Tarski [1935, 227]). The core idea is to introduce n-rowed sequences of the form $s = (f_1, \dots, f_n)$, where each f_i represents a sequence of objects of a given semantic category i. A single, unified satisfaction predicate can then be defined as a relation between n-rowed sequences and formulas of $\mathcal{L}_n$.[71] Finally, Tarski points out that the resulting satisfaction predicate for $\mathcal{L}_n$ is of order $n + 1$; it thus belongs to a metalanguage that is "essentially richer" than $\mathcal{L}_n$ (Tarski [1935, 229–230]).[72]

The motivation underlying Tarski's modified account of satisfaction is to provide a unified semantic theory for all formulas of the language $\mathcal{L}_n$. The idea of a *language-internal, stratified* theory of semantic predicates is evaded in this context in favor of a metatheoretic approach. However, in a *Postscript* to Tarski [1935], after the discussion of languages of transfinite orders, the idea of such a stratified account of truth is taken up. Here Tarski refers explicitly to footnote 48 of Gödel [1931], as quoted above, and Gödel's suggestion to embed the semantic predicates for a type-theoretic language within the higher levels of the same language. He remarks:

> Moreover a decision [concerning the Gödel sentence] can be reached within the science itself, without making use of the concepts and assumptions of the metatheory—provided, of course, that we have previously enriched the language and the logical foundations of the theory in question by the introduction of variables of higher order. (Tarski [1935, 274].)

After a further discussion of Gödel's proposed method, Tarski comes back to that point. He notes that a semantic "metatheory can be interpreted in the theory enriched by variables of higher order" (Tarski [1935, 276]).

Turning to Carnap's *Logical Syntax* again, the definition of analyticity for his type-theoretic language LII is generally considered a variant of Tarski's metatheoretic definition of truth for a pure logical language, as already mentioned. Indeed, this text contains an extensive discussion of the necessary richness of (semantic) "syntax languages", together with a discussion of numerous metalogical results closely related to Tarski's. For instance, in §60c Carnap discusses the liar paradox based on a formal version of the Diagonal Lemma. Theorem 60c.1 states the indefinability of "analytic (in S)" in language S. (Carnap [1937, 219].) And concerning the type-theoretic system

[70]Tarski argues that the notion of satisfaction for languages with formulas including expressions of different "semantical types" has to subsume a number of "semantically distinct concepts of satisfaction" for expressions of each order level (Tarski [1935, 224–225]).

[71]This presentation of sequences is close to the modern treatment of higher-order languages described above. We saw that for a type-theoretic language $\mathcal{L}_\omega$, assignment functions of the form $\alpha\colon V_\alpha \to D_\alpha$ are specified for each set of variables of a given type.

[72]See Ray [2005] and Goméz-Torrente [2004] for detailed discussions of Tarski's related indefinability theorem and his understanding of the "essential richness" of a metalanguage.

LII, Carnap holds that "'analytic in LII' is not definable in LII, but is only definable in a still richer language" (ibid.).

While Carnap is not fully explicit about what "richer modes of expression" amount to here, it is plausible that his understanding is very similar to Tarski's treatment of "essential richness", in terms of being of "higher order" than the object language in question. Thus, when defining analyticity in LII, where LII is of order ω, the syntax language in which Carnap's definition is formulated is of order $\omega + 1$. This resembles Tarski's suggestion, in the postscript to Tarski [1935], that truth definitions can also be given for languages of infinite orders, namely in languages indexed by a still higher transfinite ordinal. Most likely, the same idea also motivates Carnap's well-known claim that "everything mathematical can be formalized, but mathematics cannot be exhausted by one system, it requires an infinite series of ever richer languages" (Carnap [1937, 222]). With respect to semantic notions, such as truth or analyticity, this implies a "Tarski hierarchy" of metalanguages and typed truth predicates defined in them.[73]

However, a second approach to expressing the semantics of LII can be found in *Logical Syntax* as well. Besides the metatheoretic definition of analyticity in a separate syntax language, Carnap proposes a *stratified* definition in a *single type-theoretic language* that reflects the direct influence of Gödel's work on type theory from the same period. The underlying idea is that one can segment a language like LII into a cumulative series of "language regions" $\mathrm{II}_1, \mathrm{II}_2, \ldots, \mathrm{II}_\omega$ (of orders $1, 2, \ldots, \omega$), where LII is the union of all of them (Carnap [1937, 88]). The notion of analyticity can then be defined partially for each region in the next higher region of LII. As Carnap puts it:

> If we take as our object-language not the whole of Language II but the single concentric regions [...], then for our syntax-language we have no need to go outside the domain of LII. It is true that the concept 'analytic in II_n' is not definable for any n in II_n itself as syntax-language, but it is always definable in a more extensive region II_{n+m} (perhaps always in II_{n+1}). Hence every definition of one of the concepts 'analytic in II_n' (for the various n), and also every criterion for 'analytic in II' with respect to a particular sentence of II, is formulable in II as syntax-language. (Carnap [2002, 113].)

Here Carnap presents an early version of the hierarchy of semantic predicates for higher "language regions" of a language within the same language. This account comes very close to our account, in Section 2.2, of defining a series of typed satisfaction predicates Sat^{n+1} for formulas φ of each sublanguage $\mathcal{L}_n$ in the type-theoretic language $\mathcal{L}_\omega$.

[73]See Friedman [1999] for a detailed discussion of Carnap's conception of (semantic) metalanguages in *Logical Syntax*.

To sum up, Carnap, Tarski, and Gödel all embrace—more or less explicitly and parallel to their metatheoretic work—the idea of a stratified theory of semantic predicates for type theory in the 1930s. Moreover, their relevant proposals show a certain *continuity* with earlier "monolinguistic" approaches to semantic notions.[74] A second line of continuity connects work on type theory from before and after the often mentioned "metalogical turn" as well, although it also involves an important change. Namely, the type-theoretic languages in Carnap [1934] and Tarski [1935] are no longer meant to provide universal systems in which both theories and their metatheory can be expressed. They are now treated as object languages and, as such, strictly separated from their metalanguages. Yet there is one respect in which they are conceived of in the *same* way as before. Namely, languages like Carnap's LII are still treated as fully interpreted, i.e., as "meaningful formalisms", and not as formal in the modern sense. This point leads directly to the second interpretive issue to be discussed further by us.

4.2. Domain variation. A central question concerning the early model theory for type theory in the literature concerns the issue of domain variation for theories expressed in them. As noted earlier, this question tends to be discussed against the background of distinguishing "two logical frameworks", type theory and first-order logic. It is often claimed that the core difference between the two corresponding traditions for formalizing axiomatic theories concerns precisely the issue of domain variability. The point is sometimes put as follows: "[T]he two approaches come apart in the 1920s exactly on the issue of interpretations of the variables" (Mancosu [2010a, 5]). A standard way to understand this claim is as follows: Whereas work within the type-theoretic tradition generally assumes a fixed interpretation of the background language, work within the first-order tradition is based on the model-theoretic conception of languages. But then, first-order logic comes with a variable domain conception *for the background language*, while type theory comes with a fixed domain conception for it.[75] Moreover, the latter

[74] From a philosophical perspective, it would be interesting to see to what extent these "internal" accounts of truth in type theory are related to, or in tension with, the "universalist" conception of logic manifest in Carnap's and Tarski's earlier contributions. We leave the discussion of this question for another occasion. See Hintikka [1992], Awodey and Carus [2001], and Reck [2007] for Carnap's logical universalism in *Untersuchungen*. See Feferman [2008] for a discussion of Tarski's logical universalism in Tarski [1935].

[75] A fixed, "universal" interpretation of STT is usually considered a basic, non-negotiable assumption in the *logicist* tradition, from Frege and Russell on. As Warren Goldfarb put it in an influential article: "The ranges of the quantifiers—as we would say—are fixed in advance once and for all. The universe of discourse is always the universe, appropriately striated." Goldfarb [1979, 352]. Similar claims have recently been made concerning some of Tarski's early works. Compare again Mancosu: "It is the type-theoretic framework that in the first place decides what the class of individuals (V) is. [. . .] The type-theoretic framework comes interpreted. In particular, the quantifiers play the role of logical constants and thus [. . .] there is no reinterpretation of the quantifiers." (Mancosu [2006, 232].)

is assumed to prevent domain variation *for mathematical theories* in the type-theoretic tradition as well.

This characterization may well be correct for logic in the 1920s, at least as a broad classification. However, a closer look at logical practice in the 1930s shows that the situation is considerably more complicated during that period, in two respects. First, while both Carnap and Tarski usually work with fully interpreted type-theoretic logics in the 1930s, different techniques can be identified in their writings that allow them, nevertheless, to effectively vary the interpretation of their *background languages*. In a recent discussion of Tarski's conception of model (Mancosu [2006]), a related distinction has been drawn between two semantic conceptions of STT, a "weak" and a "strong fixed domain conception." According to the strong version, the individual quantifiers of STT always range over a fixed class of logical objects, such as Tarski's class of individuals V. According to the weak version, various higher-order languages are introduced as "specific interpreted languages" for a given axiomatic theory. The basic idea concerning the latter is this:

> Every theory comes equipped with its background theory of types and with its own interpretation of the theory of types. [...] Thus, *there is a certain flexibility in choosing what the class of individuals is that will be assumed in the background.* (Mancosu [2006, 233, emphasis added].)

For example, when Tarski uses STT for formalizing an arithmetical theory, the individual variables are stipulated to range over the class of natural numbers. When it is used for formalizing an axiom system of geometry by him, the individual variables are stipulated to range over points (lines, planes, etc.). Thus, the domain of individuals can be "interpreted" differently in different theoretical contexts, by substituting different classes of mathematical objects for it. Furthermore, no logically privileged class of individuals is assumed to exist. A similar "weak" conception for languages like LII can be identified in Carnap's work on general axiomatics.[76] Note, however, that even in the weak conception, as adopted by Tarski and Carnap at the time, the interpretation of STT remains fixed relative to a given context. In other words, once the type-theoretic background language is set up for a specific mathematical formalization, its interpretation remains fixed.

A second complication is independent, or orthogonal, to the issue highlighted by Mancosu. In Carnap's and Tarski's works from the 1930s, several further conventions can be identified that allow them to simulate the domain variability of models for *axiomatic theories* even assuming a fully interpreted type-theoretic language in the background. Recall here that, in contrast to Gödel's and Hilbert & Bernays' definitions of satisfiability, the domain of

[76]See Schiemer [2013] for details.

a model is not explicitly specified in the definitions of truth in a model in Carnap [1930], Carnap and Bachmann [1936], or Tarski [1936]. In Tarski's case, recent scholarship has shown that in his work on the "methodology of deductive sciences" he frequently uses "domain predicates", i.e., additional primitive terms whose purpose is to "specify" a particular domain of discourse for an axiomatic theory expressed in STT.[77] But as subsequent debates have also indicated, it is notoriously difficult to determine what exactly "specifying the domain" means for Tarski in this context.

In fact, in writings on semantics from the 1930s—now by all the logicians considered in this paper, not just Tarski—several *different* conceptions of the logical role of domain predicates can be identified:[78]

(1) A domain predicate $P(x)$ is added to a theory so as to "name" the range of quantification of the background language $\mathcal{L}$ in every interpretation of the theory. The individual domain D of $\mathcal{L}$ is thus identified with the extension of $P(x)$ under a given interpretation $\mathcal{M}$, i.e., $P^{\mathcal{M}} = D$.[79]

(2) An added domain predicate $P(x)$ is used to effectively relativize the fixed interpretation of the quantifiers of $\mathcal{L}$ to a subset of the intended universe D for certain purposes. More specifically, the range of quantification is restricted to a particular "domain of discourse" of an axiomatic theory expressed in $\mathcal{L}$. That theory's domain is then the extension of $P(x)$ under a suitable interpretation $\mathcal{M}$, and $P^{\mathcal{M}} \subseteq D$, while the domain of the unrestricted quantifiers is still D.[80]

  (2i) In a subversion of (2), an axiom of the form $\forall x\, P(x)$ stipulates that the extension of the domain predicate coincides with D. Therefore, as in (1), $P^{\mathcal{M}} = D$.[81]

Besides these options, recall that in the 1930s the term "model" is often defined for axiomatic theories represented in *purely* logical languages (as opposed to languages with non-logical constants). In that context, two further conventions to simulate domain variability can be distinguished:

(3) Instead of using a non-logical domain predicate $P(x)$ of $\mathcal{L}$, as above, the domain of a theory formalized in the pure language $\mathcal{L}'$ is specified

[77]Concerning this use of domain predicates by Tarski, Mancosu comments: "When a mathematical theory is formalized within this background, the domain of individuals is given at the outset by the type theory and the mathematical theory being formalized within this logical system will generally have a specific predicate that characterizes the domain of elements the theory is talking about." (Mancosu [2010a, 5].)

[78]Our aim in what follows is not to present a conclusive interpretation of Tarski's use of such predicates, but to provide an overview of different (partly conflicting) ways of understanding their uses in the formalization of axiomatic theories in the 1930s.

[79]In Goméz-Torrente [1996], such a conception is attributed to Tarski for his work in the 1930s. But compare Goméz-Torrente [2009] for a qualification of his suggestion.

[80]See Mancosu [2006], Bays [2001], and Goméz-Torrente [2009] for discussions of this understanding of domain predicates in Tarski's work.

[81]For uses of this convention in Tarski's work on categoricity, see Mancosu [2010b].

by the assignment of an extension $P^{\mathcal{M}}$ defined on D to a particular unary relation variable $X(x)$. As in (2), we then have $P^{\mathcal{M}} \subseteq D$.

(4) Sometimes no variable is used for this purpose, while a theory's domain is identified more implicitly. In particular, such domains are conceived as the (union of the) fields of all n-tuples of relations on D that are assigned to the primitive terms of an axiomatic theory T. For the implicit domain A of the theory, we thus have: $A \subseteq D$.[82]

Conventions (2), (3), and (4) are closely related to each other; they are slightly different variants of the same basic idea for specifying models of a theory in a fully interpreted background language. They all produce, in effect, relativizations in the standard model-theoretic sense.[83]

Clear uses of domain predicates in sense (2) can be found in Tarski's work on the model theory of "deductive theories" in the 1930s, for instance in his logic manual, *Einführung in die mathematische Logik und die Methodologie der Mathematik* (Tarski [1937]).[84] Tarski's treatment of the relative concepts of "satisfaction" and "truth in a" (where a is a subset of the intended interpretation of LCC) in Tarski [1935] also conforms roughly to (2); but it cannot be considered a case of relativization in the strict sense, since he does not specify a syntactic expression—a domain predicate or a class term—in his language that actually restricts the individual domain to the set a.

For our purposes, it is important to be precise about some differences between the kinds of relativization resulting in the four cases above. (1) and (2) are sometimes both characterized as effecting a relativization of the quantifiers of the type-theoretic background language (e.g., Mancosu [2010a, 9–10]). But whereas this is correct with respect to (2), it is not really what happens in (1). Let $\mathcal{M} = (D, I)$ be an interpretation of language $\mathcal{L}$ with the non-logical predicate $P(x)$. In version (2), the quantifiers of $\mathcal{L}$ range over D but can also be restricted—in terms of the relativized formulas '$\forall x\,(P(x) \to \dots)$' and '$\exists x\,(P(x) \wedge \dots)$'—to different interpretations of $P(x)$ in $\mathcal{M}$. By contrast, in version (1) the quantifiers of $\mathcal{L}$ do not range over a fixed D, but are specified to range directly over the interpretations of the domain predicate. Here quantifiers are not relativized in the usual sense, but in fact reinterpreted with every new interpretation of $P(x)$.

An example of (1), or of an approach related to (1) (but using the "variabilization" of non-logical terms), can be found in Hilbert & Bernays' *Grundlagen der Mathematik*, Volume I. After their definition of "satisfiability" for an axiomatic theory, say $\mathsf{U}(R, S)$ with two primitive terms R and S, the specification of the domain of a model is explained as follows:

[82] In Schiemer [2013], it is argued that this is characteristic for Carnap's conception of models both in *Untersuchungen* and in Carnap and Bachmann [1936].

[83] One could formulate modifications of our definitions in Section 2 along such lines. These conventions are also similar to the technique of constructing inner models for ZFC via the relativization of formulas to subsets of the universe V. See Jech [2006].

[84] See again Bays [2001] and Mancosu [2006] for further details.

> It should be noted that together with the specification of the predicates $[R, S]$ one also has to fix the domain of individuals to which the variables $x, y, \ldots$ are to refer. In a certain sense, it is introduced in the logical formula as a hidden variable. (Hilbert and Bernays [1934, 8–9].)

The stipulation of the domain of individuals in terms of an assignment to a hidden, implicit domain variable clearly implies a variable domain conception concerning the interpretations of $\mathsf{U}(R, S)$. It also implies a variable domain conception for the logical background language in use.

When made more explicit, the logical formalization of a theory along such lines would include an extra unary predicate variable that constitutes the range of individuals of the language. In fact, models are here explicitly understood as the combination of the "individual domain" of a theory and the chosen interpretations of its primitive terms. Consequently, "the satisfaction of the axioms" in them can be proved (Hilbert and Bernays [1934, 12]). Note that the same basic effect, namely the identification of the general range of quantification with the extension of a domain predicate, is achieved with convention (2.1) where $P^{\mathcal{M}}$ is axiomatically stipulated to coincide with D. However, in that case a fixed domain D of the background language is assumed, so that a significant difference remains.[85]

When using (3) or (4), a theory is expressed in the pure language $\mathcal{L}'$ with an interpretation $\mathcal{M} = (D, \mathit{Rel})$. In the case of (3), the domain of a theory relative to $\mathcal{M}$ is specified by assigning a unary relation (or set) $R \in \mathit{Rel}$ to a specific relation variable X. As in (2), D is effectively relativized to R in formulas of the form '$\forall x\,(X(x) \rightarrow \ldots)$'. This is most likely the method implicit in Tarski's account of "truth in a model" in Tarski [1936]. A more explicit example of relativization in sense (3) occurs in Carnap's *Abriss der Logistik*.[86] Finally, Carnap's *Untersuchungen zur allgemeinen Axiomatik* is best interpreted in terms of convention (4). At least in some of his examples of formalized axiomatic theories, no specific variable is reserved for the specification of the respective domains. Instead, for a theory expressed in language $\mathcal{L}'$ with interpretation $\mathcal{M}$, the range of individuals D is restricted to a set A that is the union of all the fields of relations (in *Rel*) assigned to the "primitive signs" of the theory (usually second-order relation variables).

[85]Some objections to the use of domain predicates in sense (1)—directed against Gomez-Torrente's interpretation of Tarski—are discussed in Mancosu [2006] and Mancosu [2010a]. Thus, Mancosu points out that perfectly well-formed sentences like $\exists x\, \neg P(x)$ could not be expressed in a theory containing a domain predicate $P(x)$ since, according to convention (1), they would be equivalent to the inconsistent statement $\exists x\, P(x) \wedge \neg P(x)$. When using convention (2), (3), and (4), in contrast, $\exists x\, \neg P(x)$ simply expresses the fact that there exists an object outside the particular domain of the theory in question.

[86]Thus, Carnap presents an axiomatization of Peano arithmetic with three primitive terms *nu*, *za*, *Nf* (for Zero, Number, and Successor, respectively), among others. In several axioms, the unary predicate *za* is used to relativize a universal quantifier to the set of natural numbers exactly in sense 3. See Carnap [1929, 74–75].

§5. **Conclusion.** The main objective of this paper was to provide a survey of several contributions to formal semantics in the 1930s, with particular focus on those presented in type theory. We showed that the corresponding early definitions of formal models and model-theoretic truth anticipated central features of our modern understanding of these notions. Work on semantics during this phase was clearly formative for the subsequent consolidation of model-theoretic semantics in the 1950s. As also became evident, however, the earlier notions discussed by us are not identical to the modern ones. Characteristic aspects of the model-theoretic accounts of truth and validity are still missing—and not only in the type-theoretic approaches of Tarski and Carnap, but also in the definitions of Gödel and Hilbert based on first-order logic.

While some central features of the later model-theoretic approach are absent in these contributions from the 1930s, we illustrated, furthermore, that logicians used a variety of ways to simulate their effects. In particular, we surveyed several different conventions that allow for the simulation of domain variability for axiomatic theories expressed in a type-theoretic language, even if the latter is assumed to have a fixed interpretation. We also established that a proper object-/metalanguage distinction was often recast, or prefigured, in terms of a stratified account of truth defined within a single type-theoretic system. Together these techniques yield results that are not identical but commensurable with modern model theory.

Overall, this suggests a reevaluation of our understanding of the "semantic turn" in mathematical logic, and in particular, of the transition from type theory to the modern model-theoretic conception of logic. The usual emphasis on two radically different, incompatible "logical frameworks"—type-theory and first-order logic—needs to be at least qualified, in some respects also revised more substantively. Not only were significant advances towards model-theoretic semantics made within the type-theoretic tradition, e.g., by Carnap and the early Tarski; the view that proponents of the first-order tradition, like Gödel and Hilbert, fully embraced the current model-theoretic conception of logic already in the 1930s is also clearly misleading.

As should have become evident, there are in fact striking similarities between the various treatments of notions like satisfaction, truth, and validity considered in this paper, across the border of the two logical camps. This illustrates that the evolution of model-theoretic logic was a more multifaceted phenomenon than is often assumed. Its source was not only the first-order tradition. Indeed, there was no single source, or discrete "historical birth", of model theory. Rather, there was a continuous development with contributions from all logical schools active at the time, including the type-theoretic tradition. Our survey was restricted to activities in the 1930s.

A similar study of later developments, up to the 1950s, still remains to be done.[87]

REFERENCES

P. B. ANDREWS [2002], ***An introduction to mathematical logic and type theory: To truth through proof***, Kluwer, Dordrecht.

S. AWODEY AND A. CARUS [2001], *Carnap, completeness, and categoricity*: *The* Gabelbarkeitssatz *of* 1928, ***Erkenntnis***, vol. 54, pp. 145–172.

S. AWODEY AND A. CARUS [2007], *Carnap's dream*: *Gödel, Wittgenstein and the Logical Syntax*, ***Synthese***, vol. 159, pp. 23–45.

S. AWODEY AND A. CARUS [2009], *From Wittgenstein's prison to the boundless ocean*: *Carnap's dream of logical syntax*, ***Carnap's logical syntax of language*** (P. Wagner, editor), Palgrave, London, pp. 79–108.

S. AWODEY AND E. RECK [2002], *Completeness and categoricity, part* 1: *Nineteenth-century axiomatics to twentieth-century metalogic*, ***History and Philosophy of Logic***, vol. 23, pp. 1–30.

C. BADESA [2004], ***The birth of model theory. Löwenheim's theorem in the frame of the theory of relatives***, Princeton University Press, Princeton.

T. BAYS [2001], *On Tarski on models*, ***The Journal of Symbolic Logic***, vol. 66, pp. 1701–1726.

T. BONK AND J. MOSTERIN [2000], *Editor's introduction*, ***Carnap, R., Untersuchungen zur Allgemeinen Axiomatik,*** Carnap [2000], pp. 1–54.

R. CARNAP [1929], ***Abriss der Logistik***, Springer, Wien.

R. CARNAP [1930], *Bericht über Untersuchungen zur allgemeinen Axiomatik*, ***Erkenntnis***, vol. 1, pp. 303–307.

R. CARNAP [1934], ***Logische Syntax der Sprache***, Springer, Wien, English translation: Carnap [1937].

R. CARNAP [1935], *Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik*, ***Monatshefte für Mathematik und Physik***, vol. 42, pp. 163–190.

R. CARNAP [1937], ***The logical syntax of language***, Open Court, La Salle, republished in 2012.

R. CARNAP [2000], ***Untersuchungen zur Allgemeinen Axiomatik***, (T. Bonk and J. Mosterin, editors), Wissenschaftliche Buchgesellschaft, Darmstadt.

R. CARNAP AND F. BACHMANN [1936], *Über Extremalaxiome*, ***Erkenntnis***, vol. 6, pp. 166–188, English translation: *On extremal axioms*, ***History and Philosophy of Logic***, vol. 2 (1981), pp. 67–85.

A. CHURCH [1976], *Comparison of Russell's resolution of the semantical antinomies with that of Tarski*, ***The Journal of Symbolic Logic***, vol. 41, pp. 747–760.

A. COFFA [1991], ***The semantic tradition from Kant to Carnap: To the Vienna Station***, Cambridge University Press, Cambridge.

R. CREATH [1990], *The unimportance of semantics*, ***Proceedings of the 1990 biennial meeting of the Philosophy of Science Association*** (A. Fine, M. Forbes, and L. Wessels, editors), vol. 2, The University of Chicago Press, pp. 405–416.

W. DEMOPOULOS [1994], *Frege, Hilbert, and the conceptual structure of model theory*, ***History and Philosophy of Logic***, vol. 15, pp. 211–225.

S. FEFERMAN [2008], *Tarski's conceptual analysis of semantical notions*, ***New essays on Tarski and philosophy*** (D. Patterson, editor), Oxford University Press, Oxford, pp. 72–93.

[87] We would like to thank the editors and an anonymous referee for helpful comments on an earlier version of this paper. We are also grateful to the Munich Center for Mathematical Philosophy, especially its founder and co-director Hannes Leitgeb, for facilitating the collaboration between the two authors.

J. FERREIRÓS [2001], *The road to modern logic—an interpretation*, this BULLETIN, vol. 7, pp. 441–484.

J. FERREIRÓS [2007], ***Labyrinth of thought: A history of set theory and its role in modern mathematics***, 2nd (revised) ed., Birkhäuser, Basel.

M. FRIEDMAN [1999], ***Reconsidering logical positivism***, Cambridge University Press, Cambridge.

K. GÖDEL [1929], ***Über die Vollständigkeit des Logikkalküls***, Ph.D. thesis, University of Vienna, English translation: Gödel [1986, 60–101].

K. GÖDEL [1930], *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, ***Monatshefte für Mathematik und Physik***, English translation: Gödel [1986, 102–123].

K. GÖDEL [1931], *Über formal unentscheidbare Sätze der* Prinicpia Mathematica *und verwandter Systeme I*, ***Monatshelfte für Mathematik und Physik***, English translation: Gödel [1986, 144–195].

K. GÖDEL [1986], ***Collected Works, Volume I: Publications 1929–1936***, (S. Feferman et al., editors), Oxford University Press, Oxford.

K. GÖDEL [2003], ***Collected Works, Volume IV: Selected Correspondence, A–G***, (S. Feferman et al., editors), Oxford University Press, Oxford.

W. GOLDFARB [1979], *Logic in the twenties*: *The nature of the quantifier*, ***The Journal of Symbolic Logic***, vol. 44, pp. 351–368.

W. GOLDFARB [2005], *On Gödel's way in*: *The influence of Rudolf Carnap*, this BULLETIN, vol. 11, pp. 185–193.

M. GOMÉZ-TORRENTE [1996], *Tarski on logical consequence*, ***Notre Dame Journal of Formal Logic***, vol. 37, pp. 125–151.

M. GOMÉZ-TORRENTE [2004], *The indefinability of truth in the "Wahrheitsbegriff"*, ***Annals of Pure and Applied Logic***, vol. 126, pp. 27–37.

M. GOMÉZ-TORRENTE [2009], *Rereading Tarski on logical consequence*, ***Review of Symbolic Logic***, vol. 2, pp. 249–297.

P. HÁJEK AND P. PUDLÁK [1993], ***Metamathematics of first-order arithmetic***, Springer, Berlin.

D. HILBERT AND W. ACKERMANN [1928], ***Grundzüge der theoretischen Logik***, Springer, Berlin, 2nd ed., 1938.

D. HILBERT AND B. BERNAYS [1934], ***Grundlagen der Mathematik***, vol. 1, Springer, Berlin.

D. HILBERT AND B. BERNAYS [1939], ***Grundlagen der Mathematik***, vol. 2, Springer, Berlin.

J. HINTIKKA [1988], *On the development of the model-theoretic viewpoint in logical theory*, ***Synthese***, vol. 77, pp. 1–36.

J. HINTIKKA [1992], *Carnap's work in the foundations of logic and mathematics in a historical perspective*, ***Synthese***, vol. 93, pp. 167–189.

W. HODGES [1986], *Truth in a structure*, ***Proceedings of the Aristotelian Society***, vol. 86, pp. 135–151.

W. HODGES [2009], *Set theory, model theory, and computability theory*, ***The development of modern logic*** (L. Haaparanta, editor), Oxford University Press, Oxford, pp. 471–498.

I. JANÉ [2006], *What is Tarski's* common *concept of consequence*?, this BULLETIN, vol. 12, pp. 1–42.

T. JECH [2006], ***Set theory***, 3rd ed., Springer, Berlin.

A. KANAMORI [2007], *Gödel and set theory*, this BULLETIN, vol. 13, pp. 153–188.

J. KEMENY [1956a], *A new approach to semantics. Part* 1, ***The Journal of Symbolic Logic***, vol. 22, pp. 1–27.

J. KEMENY [1956b], *A new approach to semantics. Part* 2, ***The Journal of Symbolic Logic***, vol. 22, pp. 149–161.

O. LINNEBO AND A. RAYO [2012], *Hierarchies ontological and ideological*, ***Mind***, vol. 121, pp. 269–308.

P. MANCOSU [2006], *Tarski on models and logical consequence*, ***The architecture of modern mathematics: Essays in history and philosophy*** (J. Ferreirós and J. Gray, editors), Oxford University Press, Oxford, pp. 209–237.

P. MANCOSU [2010a], *Fixed- versus variable-domain interpretations of Tarski's account of logical consequence*, ***Philosophy Compass***, vol. 5, pp. 745–759.

P. MANCOSU [2010b], ***The adventure of reason: Interplay between philosophy of mathematics and mathematical logic, 1900–1940***, Oxford University Press, Oxford.

P. MANCOSU, R. ZACH, AND C. BADESA [2009], *The development of mathematical logic from Russell to Tarski*, ***The development of modern logic*** (L. Haaparanta, editor), Oxford University Press, Oxford, pp. 318–470.

R. MURAWSKI [1999], ***Recursive functions and metamathematics***, Kluwer, Dordrecht.

W. V. QUINE [1986], ***Philosophy of logic***, 2nd ed., Harvard University Press.

G. RAY [2005], *On the matter of essential richness*, ***Journal of Philosophical Logic***, vol. 34, pp. 433–457.

E. RECK [2004], *From Frege and Russell to Carnap*: *Logic and logicism in the* 1920*s*, ***Carnap brought home: The view from Jena*** (S. Awodey and C. Klein, editors), Open Court, La Salle, pp. 151–180.

E. RECK [2007], *Carnap and modern logic*, ***The Cambridge companion to Carnap*** (M. Friedman and R. Creath, editors), Cambridge University Press, Cambridge, pp. 176–199.

E. RECK [2013], *Developments in logic*: *Carnap, Gödel, and Tarski*, ***The Oxford handbook of the history of analytic philosophy*** (M. Beaney, editor), Oxford University Press, Oxford, pp. 546–571.

G. SCHIEMER [2012a], *Carnap on extremal axioms, "completeness of the models," and categoricity*, ***The Review of Symbolic Logic***, vol. 5, no. 4, pp. 613–641.

G. SCHIEMER [2012b], *Carnap's* Untersuchungen: *logicism, formal axiomatics, and metatheory*, ***Rudolf Carnap and the legacy of logical empiricism*** (R. Creath, editor), Springer, Berlin, pp. 13–35.

G. SCHIEMER [2013], *Carnap's early semantics*, ***Erkenntnis***, vol. 78, no. 3, pp. 487–522.

S. SHAPIRO [1991], ***Foundations without foundationalism: A case for second-order logic***, Oxford University Press, Oxford.

J. SHOENFIELD [1967], ***Mathematical logic***, Addison-Wesley, Reading, Mass..

A. TARSKI [1931], *Sur les ensembles définissables de nombres réels, I*, ***Fundamenta Mathematica***, vol. 7, pp. 219–239, English translation: Tarski [1983, 110–142].

A. TARSKI [1935], *Der Wahrheitsbegriff in den formalisierten Sprachen*, ***Studia Philosophica***, vol. 1, pp. 261–405, English translation: Tarski [1983, 152–267].

A. TARSKI [1936], *Über den Begriff der logischen Folgerung*, ***Actes du Congrés International de Philosophie Scientifique***, vol. 7, pp. 1–11, English translation: Tarski [1983, 409–420].

A. TARSKI [1937], ***Einführung in die mathematische Logik und die Methodologie der Mathematik***, Springer, Vienna.

A. TARSKI [1954], *Contributions to the theory of models I*, ***Indagationes Mathematicae***, vol. 16, pp. 572–581.

A. TARSKI [1983], ***Logic, semantics, metamathematics***, 2nd ed., Hackett, Indianapolis.

A. TARSKI AND R. L. VAUGHT [1957], *Arithmetical extensions of relational systems*, ***Compositio Mathematica***, vol. 13, pp. 81–102.

MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY (MCMP)
LUDWIG-MAXIMILIANS UNIVERSITÄT MÜNCHEN, D-80539 MÜNCHEN, GERMANY
E-mail: georg.schiemer@univie.ac.at

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF CALIFORNIA AT RIVERSIDE, RIVERSIDE, CA 92521, USA
E-mail: erich.reck@ucr.edu