THE REVIEW OF SYMBOLIC LOGIC
Volume 8, Number 2, June 2015

THE EXPRESSIVE POWER OF TRUTH

MARTIN FISCHER
MCMP, LMU München
LEON HORSTEN
Department of Philosophy, University of Bristol

Abstract. There are two perspectives from which formal theories can be viewed. On the one hand, one can take a theory to be about some privileged models. On the other hand, one can take all models of a theory to be on a par. In contrast with what is usually done in philosophical debates, we adopt the latter viewpoint. Suppose that from this perspective we want to add an adequate truth predicate to a background theory. Then on the one hand the truth theory ought to be semantically conservative over the background theory. At the same time, it is generally recognised that the central function of a truth predicate is an expressive one. A truth predicate ought to allow us to express propositions that we could not express before. In this article we argue that there are indeed natural truth theories which satisfy both the demand of semantical conservativeness and the demand of adequately extending the expressive power of our language.

§1. Introduction. It is generally agreed that a truth predicate has an expressive function. A core function of a truth predicate is to increase our conceptual repertoire. A truth predicate widens the class of thoughts that we can express.

According to the axiomatic tradition, the meaning of a truth predicate is spelled out at least in part by postulating basic principles by which it is governed. In order to adequately represent the expressive function of truth a minimally satisfactory theory of truth should not be conceptually reducible to the background theory: it should not be possible to simulate the truth predicate in the background theory.

Most existing axiomatic theories of truth postulating an irreducible truth predicate result in a theory of truth that is non-conservative over its background theory. Indeed, it is often thought that non-conservativeness can be a virtue of truth theories. It is taken to demonstrate how a truth predicate can help us to zoom in on the intended interpretation of our theory. This is then interpreted as a manifestation of the substantiality of truth, and used as an argument against deflationism.

Yet conceptual irreducibility does not entail non-conservativeness. In certain contexts it can even be reasonable to insist on a truth predicate that is both conceptually irreducible and conservative. There are contexts where one is reluctant to privilege one model over another, and where one does not want a theory of truth to exclude models for the original language. In particular, this is the case in the single mathematical field where truth predicates play a major role, viz. *model theory*, and in uses of model theory in proof theory.

Received: June 6, 2014.

doi:10.1017/S1755020314000392

In this sense, our aim is similar to that of Tarski (1935) in his 'Der Wahrheitsbegriff in den formalisierten Sprachen'. Tarski did not aim to give a theory of truth for natural language: he thought that this concept is incoherent. Instead, he was concerned with establishing beyond reasonable doubt that the uses of truth predicates in metamathematics are legitimate. On the other hand, Tarski's aims were of course not entirely the same as ours. He wanted to give a *definition* of truth in formalised languages, and that is of course not our objective.[1]

In this article, we will explore the role that a truth predicate as an irreducible but conservative notion plays in model theory. For this purpose we will make use of a primitive truth predicate in contrast to the usual set theoretic definition of truth in model theory, because an axiomatic approach to truth allows for a minimal account of formalising a model theoretic notion of truth. We propose that a truth theory that is conservative over the base theory but non-interpretable in it, is sufficiently natural to capture key uses that are made of a truth predicate in model theory.

There are additional reasons for being interested in the way in which a truth predicate functions in the clean context of model theory. First, model theory is an area in which truth predicates are used as intended. Second, as a mathematical discipline, model theory provides an environment in which several uses of truth can be compared objectively. Third, the expressive powers of mathematical systems are measurable, thus allowing us to investigate the notion of increase of expressive power.

This article is structured as follows.

First, the notion of expressiveness is investigated: an explication is given of what it means for a formal language to be at least as expressive as another language. We will argue that a question of expressive strength of a language should always be posed relative to a theory formulated in the language. In our investigation, we will restrict our attention to extensional aspects of meaning, and to the situation where one language is an expansion of another language. Thus we fall short of giving a general account of expressiveness. Nevertheless, our theory is sufficiently general to allow us to judge whether a concept of truth adds genuine expressive strength to a language (relative to a theory).

Second, we describe an attractive axiomatic theory of truth that is semantically conservative over its background theory, but is not relatively interpretable in it. Since it is not interpretable in its background theory, it is not conceptually reducible to it. Therefore it is a candidate for capturing the expressive power of truth.

Third, we show that this truth theory does capture the uses that are made of truth in model theory and proof theory in a natural way. For this purpose, we discuss some of the principal uses that are made of the notion of truth in model theory and proof theory. First, we describe the connection between the notion of truth and problems concerning the finite axiomatisability of theories. Second, we will explicate the role of truth in proofs of a paradigmatic model-theoretic theorem: Gödel's completeness theorem. Third, we discuss the phenomenon of speed-up as a sign of the expressive power of truth. And fourth, we will investigate the role of the concept of truth in proving reflection principles. It will emerge that the formal theory of truth that we propose neatly covers all these cases.

§2. The expressive power of the truth predicate. One of the main themes of this paper is the conviction that the expressive function of a truth predicate is central to an

[1] Compare Hodges (2008).

understanding of truth. This amounts to saying that the truth predicate allows us to express things we cannot express otherwise. So in some sense the truth predicate has an irreducible expressive function.[2]

The expressive function of the truth predicate is often illustrated by examples such as the Law of Excluded Middle for a language $\mathcal{L}$:

For all sentences ϕ of $\mathcal{L}$, either ϕ is true or ϕ is not true.

As there are infinitely many instances of this law, we could not have expressed the same proposition by asserting all of them one by one. Such ordinary uses of truth are ubiquitous in mathematical logic. For instance, when a mathematical logician glosses her proposition by saying

(IC) "Every PA independent Π_1-sentence is true",

her use of the concept of truth is in our final analysis of the same kind as when she says "Everything Judy said yesterday is true". The truth predicate is indispensable in these cases, and in this sense the truth predicate is irreducible.[3]

Our reasoning needs some explaining. Usually model theorists would take the notion of truth for arithmetic to be the notion of truth in the standard model for arithmetic. This is reasonable in a context with a strong background theory such as set theory. However there are viable alternatives. In the example (IC) the complexity of the sentences is restricted so that we could use a definable restricted truth predicate for a formalisation. In cases where the complexity of the sentences is not restricted we have to make use of a primitive notion of truth or a stronger background theory. If we want to address the question of a minimal framework sufficient to formalize the claim (IC) then we opt for an axiomatic theory of truth. And in this context the model theoretic claim (IC) can be understood in terms of ordinary use of truth. In the end all the formalisations refer back to some intuitions a truth predicate has to satisfy. We can judge whether a definition of truth is adequate only by considering criteria of adequacy. However, sometimes those criteria can be directly understood as axioms, and therefore we consider the axioms to be basic.

The expressive function of the truth predicate is widely recognised. Deflationist accounts of truth even take the expressive role of truth to be central. Some deflationists go further than this and claim that the expressive function is the sole function of the truth predicate.[4] In this article we do not sign up to any such further constraints that some deflationists want to impose on theories of truth. We wish to investigate what truth theories are minimally committed to if we take as our single desideratum that they capture the expressive function of the concept of truth.

Expressiveness is in need of explication in several ways. One question concerns the objects to which we attribute expressiveness. Confining ourselves to formal settings, there are at least three options: Languages, logics or theories.

---

[2] An example of such a view is Quine (1970, chap. 1). Quine emphasised the irreducibility of the notion of truth in cases where one wants to ascertain an infinite lot of sentences.

[3] This is not to say that we could not introduce devices such as propositional quantification to express these 'infinitary' statements. Usually such devices are connected with commitments that we could avoid if we would have a truth predicate instead. With a truth predicate we can stay within the confines of extensional first-order logic. We merely have to specify rules for the truth predicate for it to be able to function as an expressive device.

[4] See, for example, Horwich (1998).

It is not clear whether there is a primary use, although it seems more natural to apply the concept of expressiveness to languages and to logics. To attribute expressive power to formal theories almost seems to be an abuse of language. Theories have deductive power, which has to be separated from expressive power; only in a somewhat derivative sense is it admissible to apply the notion of expressiveness to theories. Yet in the final analysis expressiveness is connected to logics, languages, and theories, as they are intertwined in a way that makes a strict separation impossible. The aim of this section is to try to make sense of expressiveness for extensional first-order languages with respect to a certain class of intended models.

Expressiveness is usually understood as a semantical notion. So if we consider logics or languages, then we take them already to be equipped with a certain semantics. Model-theoretic semantics captures some of the extensional aspects of meaning, and for our modest purposes this is sufficient. So we consider formal languages equipped with a model-theoretic semantics. In other words, we consider languages as coming with a class of intended models. And as we have already announced, we will concentrate on mathematical languages. Many such languages have a large class of intended interpretations—think of the language of group theory, for instance. Other languages are often taken to have (up to isomorphism, perhaps) only one intended interpretation: the language of arithmetic is a case in point.

One indicator of expressive strength that appears to be especially suitable for languages with a unique intended model is given by the definable class of subsets of the domain of discourse of the intended interpretation. A language $\mathcal{L}$ is then said to be at least as expressive as $\mathcal{L}'$ if and only if in $\mathcal{L}$ we can define at least all the subsets of the domain of the intended structure that are definable in $\mathcal{L}'$. This strategy seems useful in the case of arithmetic, and since we are working in a setting where we identify syntactical objects with numbers, it seems also an attractive option in our situation.

Nevertheless, there are several reasons for not following this explication. First, this strategy is not very generally applicable: it can be applied only if the two languages have the same unique intended model. The intended model or standard model of arithmetic $\mathcal{N}$ is often taken to be well understood and the definable subsets of ω are subject of deep logical investigations. But this situation does not generally obtain.

Second, talk of the standard model of arithmetic is not without its problems. We cannot fix the standard model by means of a first-order theory; in order to give a categorical theory of the standard model, we have to introduce stronger means such as standard second-order semantics. For our purposes we would prefer a criterion that avoids commitments such as those introduced by standard second order logic. If we already had a notion of standard model then we could also make use of the notion of truth in the standard model, which would not allow for an analysis of how much proof-theoretic strength we need to be able to capture relevant uses of the model theoretic notion of truth. We will argue that there is an alternative criterion, which is attractive and avoids those commitments.[5]

In the general model theory, all models are usually taken as being on a par: no single model is more ‘intended’ than another. This attitude is appropriate for so-called ‘algebraic’

[5] Doubts about the usefulness of the talk of the standard model of arithmetic have also been expressed by metamathematicians such as Bovykin. Bovykin’s idea of arithmetical splitting, i.e. absolute unprovable arithmetical statements, stands in sharp contrast to the conception of the standard model.

mathematical theories.[6] From that perspective, the definition of expressiveness takes another form: one looks at classes of models characterisable by sentences of the language. This is in line with the conception that the meaning of a sentence is the class of models satisfying the sentence, rather than one intended interpretation. The general idea behind the explication is this: a language $\mathcal{L}'$ is at least as expressive as a language $\mathcal{L}$ if and only if for every sentence φ in $\mathcal{L}$ there is a sentence ψ in $\mathcal{L}'$, such that ψ characterizes a class of models which is equivalent, in the relevant sense, to the class of models characterized by φ.

Recall that for a logic L and a signature τ:

DEFINITION 2.1. *A class of τ-structures κ is an elementary class in L, in symbols $\kappa \in EC_L$, iff there is a $\varphi \in \mathcal{L}[\tau]$ such that $\kappa = \mathrm{MOD}^{\tau}_{L}(\varphi)$, where $\mathrm{MOD}^{\tau}_{L}(\varphi)$ is the class of L-models of φ.*[7]

In the field of model-theoretic logics an explication of expressive power of logics based on elementary classes has been fruitful.[8] Let L_1, L_2 be two (abstract model-theoretic) logics, i.e., a specification of the syntax and a model-theoretic semantics. L_2 is at least as expressive as L_1 if and only if every elementary class in L_1 is elementary in L_2. This explication is adequate to order logics with respect to their expressive power. It has, for example, the consequence that second-order logic is at least as expressive as first-order logic. It is also used in Lindström's characterisation of first-order logics as the strongest logic that has the properties of compactness and Löwenheim-Skolem.[9]

Whereas the comparison of different logics is very interesting on its own, we want to focus on theories of truth formulated as first-order theories.[10] There are at least three problems with transferring the model-theoretic criterion for expressiveness of logics to first-order languages. Firstly, we are now talking about specific languages with fixed signatures and we need a way to compare them. Secondly, languages considered only as given by their signature and inductive definition of well-formed formula without a class of intended interpretations or a theory would trivialise the criterion. Third, we do not want an explanation that reduces to a comparison of deductive power of theories formulated in these languages.

In order to answer the first problem we consider a simplified case where one language is a subset of the other language. In this case it is obvious that the second language is at least as strong as the first. Moreover in order to be able to classify languages as properly stronger

---

[6] For the distinction between algebraic mathematical theories (such as group theory) and non-algebraic mathematical theories (such as real analysis), see Shapiro (1997, pp. 40–41).

[7] Sometimes two notions of elementary classes are distinguished. In addition to the given definition we have that a class of τ-structures κ is elementary in L in the wider sense if there is a set of sentences Δ in $L[\tau]$, such that $\kappa = \mathrm{MOD}^{\tau}_{L}(\Delta)$.

[8] Barwise & Feferman (1985).

[9] Barwise & Feferman (1985).

[10] There are multiple reasons for this. First, there is the pragmatic reason that the best investigated axiomatic truth theories are first-order theories. Second, first-order logic seems sufficient for many purposes. For example set theory is a tool available that allows to formalise most of mathematics and seems therefore sufficient for most purposes. Third, a background logic that is too strong would conceal the differences of the axiomatic theories of truth we want to analyse. For our minimalist aims it seems therefore natural to stick to first-order logic. This is not supposed to be a convincing argument for a first-order thesis that argues for first-order logic as the only viable logic, but rather a justification of our focus. For an argument against second-order logic as the underlying logic the reader may consult Väänänen (2001).

than others it is sufficient to look at model expansions and model reducts to compare classes of models with two different signatures. For a more general explication we would need a way to compare languages with different signatures, which could be accomplished by structure preserving translations and a corresponding function on the classes of models.[11] These translations would allow us to abstract away from any specific signature and talk about classes of languages.

This solves our first problem of comparing models of different signatures. But if we do not restrict the possible interpretations, then a problem of trivialisation arises.[12] So far we did not consider a restriction of the interpretation of the vocabulary under consideration. However we want to compare the expressive power of a language without a truth predicate with a language containing a truth predicate and therefore we have to restrict the class of models in an appropriate way. This observation alone shows that a strict separation between languages and theories is not adequate for an explication of expressiveness. To avoid this trivialization, we will focus on expressiveness of a language with respect to a fixed context, where we think of a class of (intended) structures as fixing the context. So we think of a language $\mathcal{L}$ and the class of all $\mathcal{L}$-models, $\mathrm{MOD}^{\mathcal{L}}$, restricted to a class of intended models $\mathrm{IM}_{\mathcal{L}}$, where $\mathrm{IM}_{\mathcal{L}} \subseteq \mathrm{MOD}^{\mathcal{L}}$. In the interest of readability we have simplified the expression by dropping the subscript for the logic, as we only consider first-order logic. Moreover we replaced the superscript τ for the signature by $\mathcal{L}$ in order to directly refer to the language.

In most cases we will think of an axiomatisable theory as fixing the context. Usually we are in the context of a specific subject, such as arithmetic for example. We consider the language $\mathcal{L}_A$ with signature $\tau_A = \{0, 1, S, +, \cdot\}$. Furthermore we think of the interpretation of the vocabulary as fixed to some extent. We take as a minimal criterion for a language to be an arithmetical language that the axioms of Robinson Arithmetic (Q) are satisfied. For a language of truth based on an arithmetical language we require it to satisfy at least the Tarski-biconditionals for arithmetical sentences in addition to Q.

If we relativize the languages $\mathcal{L}_1$ and $\mathcal{L}_2$ to the classes of intended models given by the theories T_1 and T_2, respectively, we have to make sure that we can recover every model of T_1 as a reduct of a model of T_2. For this we will appeal to the notion of a *semantically conservative extension*,[13] which is based on the concept of model expansion:[14]

DEFINITION 2.2. *A theory T_2 is a semantically conservative extension of T_1 if and only if every model of T_1 can be expanded to a model of T_2.*

If the classes of intended structures are given by theories we will say that:

DEFINITION 2.3. *Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Then $\mathcal{L}_2$ is expressively stronger with respect to T_2 than $\mathcal{L}_1$ with respect to T_1 (in symbols: $\mathcal{L}_1(T_1) < \mathcal{L}_2(T_2)$) iff there is a class of models κ which is*

---

[11] The translation should probably satisfy more criteria, such as preserving the vocabulary that is part of both signatures. In terms of category theory, the corresponding function on models is a contravariant functor: see Visser (2004).

[12] For example if we state the criterion as follows: Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$. $\mathcal{L}_2$ is expressively stronger than $\mathcal{L}_1$ if and only if there is an elementary class κ of $\mathcal{L}_2$ such that $\kappa' = \{\mathfrak{M}\restriction\mathcal{L}_1 \mid \mathfrak{M} \in \kappa\}$ is not elementary in $\mathcal{L}_1$. Of course this definition is only useful if $\mathcal{L}_2$ is an extension of $\mathcal{L}_1$. But that is the only situation that interests us in this article, where $\mathcal{L}_2$ is obtained by adding a truth predicate to $\mathcal{L}_1$.

[13] The distinction between syntactical and semantical conservativeness is implicit in Craig & Vaught (1958, p. 292).

[14] An expansion of a model $\mathfrak{M}$ adds interpretations for new vocabulary but the domain is preserved.

elementary in $\mathcal{L}_2$ with respect to T_2, but for which the class κ' of reducts is not elementary in $\mathcal{L}_1$ with respect to T_1.[15]

Due to the restriction to language expansions and semantically conservative extensions, this explication of expressive force indeed only has a limited scope. But it does cover all the cases that are of interest to us in this article. For instance, this explication entails that the language of class theory with respect to the class theoretic version of Kripke-Platek set theory is expressively stronger than the language of set theory with respect to KP, and the language of second-order arithmetic with respect to RCA_0 is expressively stronger than the language of first-order arithmetic with respect to $\mathrm{I}\Sigma_1$. In this article, we are only interested in certain specific cases of relative expressive power. We want to discuss the specific expressive function that a truth predicate can and does play as part of an axiomatic theory of truth. We will make use of our explication in Section 5 to show that the language of arithmetic expanded by a truth predicate is expressively stronger than the language of arithmetic itself. There are facts about the natural numbers that we cannot express in the language $\mathcal{L}_A$, no matter which theory we relativise this language to: for example, there is no sentence of arithmetic that will restrict the class of models to the singleton of the standard model, $\{\mathfrak{N}\}$ (or its isomorphism type). But there is no first-order language expansion that is expressive enough to express this. Up to isomorphism, $\{\mathfrak{N}\}$ is an EC in standard second-order logic but not in first-order. But also the class of PA-models is not elementary in $\mathcal{L}_A$, for there is no *single* sentence that has this class as its models. We will show that in this case we can expand the language by a truth predicate in such a way that the class of PA-models becomes elementary. But before doing this we want to motivate our restriction to conservative extensions.

§3. Conservativeness. Some philosophers have suggested that an adequate deflationist theory of truth should be conservative over the background theory.[16] They have claimed that this requirement follows from the fact that truth is insubstantial.[17] But all deflationists hold that truth has an expressive function. So one may wonder whether the demands of expressiveness and conservativeness are compatible, given our preferred explication of expressiveness of truth notions. After all, we take expressiveness to be intimately connected to deductive strength.

The previous section shows that there are at least two notions of conservativeness relevant to our discussion. Let S and T be theories, i.e. deductively closed sets of formulas:

1. The common proof-theoretic notion of T being a (syntactically) conservative extension of S if and only if $S \subseteq T$ and for all φ in the language of S, if T proves φ, then already S proves φ;
2. The model-theoretic notion of T being a (semantically) conservative extension of S if and only if all the models of S can be expanded to models of T.

These two notions are not equivalent. The semantical notion is stronger in the sense that every semantically conservative extension is also a syntactically conservative extension.

---

15 A more general and formal explication is given in Section 11.

16 See Horsten (1995), Ketland (1999), and Shapiro (1998).

17 For example Ketland says: '*if* truth is non-substantial—as deflationists claim—*then* the theory of truth should be *conservative*' (Ketland, 1999, p. 79). However this inference has been challenged, for example by Halbach (2014) and Horsten (2011, chap. 7).

The converse is not generally the case as is shown by the theory $\mathrm{CT}\restriction$: the Tarskian compositional theory of truth with induction restricted to the truth-free background language of arithmetic.[18] $\mathrm{CT}\restriction$ is a syntactically conservative extension of PA but $\mathrm{CT}\restriction$ is not semantically conservative, because only recursively saturated non-standard models can be expanded to models of $\mathrm{CT}\restriction$.[19]

In the literature on deflationism it is not always made explicit which notion of conservativeness is considered. The distinction between the two notions of conservativeness depends on what one takes a theory to be. If one accepts the picture of theories determining the range of interpretations as a class of models, as is generally done in the so-called semantic view of scientific theories,[20] then the semantical conservativeness is a better motivated explication of non-substantiality.[21]

Shapiro motivates the deflationist claim of conservative truth with the example of Karl, which suggests a reading of the conservativeness claim involving the notion of semantical conservativeness.[22] A semantically non-conservative theory of truth is a theory of truth that rules out some of the models of the background theory. A semantically conservative notion of truth, on the other hand, leaves all the possible interpretations of the background theory intact.

If we take the perspective of one intended interpretation of arithmetic, and the concomitant definition of expressiveness in terms of definable subsets, then of course no possibilities are eliminated by non-conservative truth theories. All models save the intended one are not what the theory is really about. So if some of them are eliminated by the truth theory: no matter! But if we take the general model theoretic (or algebraic) perspective and treat all models as being on a par, then a truth theory for a background theory should not eliminate possibilities, for then it would not be a truth theory *for* the background theory (conceived of as a class of models). The general model theoretic perspective is the one that is adopted in mathematical logic, especially in model theory and in the proof-theoretic analysis of theorems of model theory. Hence the preference in those areas for truth theories that are semantically conservative. In this article, we discuss truth theories from this model-theoretic or algebraic perspective.

All this does not entail, of course, that non-conservative truth theories are mistaken or devoid of interest. It just means that from the general model theoretic perspective that we are adopting, they are not appropriate. They are appropriate only from the perspective where one does not consider any model of a given background theory as being as good as any other.

§4. A minimally adequate theory of truth. In this section we present the truth theory PT^-. This theory was first formulated and investigated in Fischer (2009). We will argue that it satisfies the expressive uses of truth in the minimal (conservative) way that is required from the algebraic perspective that we are adopting in this article.

The theory PT^- is based on a theory of syntax, which we will take to be PA. This is not uncommon, and there is rhyme and reason to it. On the one hand we know that sequential

---

18 For a precise definition of $\mathrm{CT}\restriction$ see Halbach (2014).

19 This follows by a theorem of Lachlan: see Kaye (1991, p. 228).

20 See for example Suppe (1977).

21 Thanks to Volker Halbach for pointing this out to us.

22 See Shapiro (1998), 487f.

arithmetical theories allow for an arithmetisation of parts of metamathematics including theories of syntax. On the other hand PA is a very natural theory of arithmetic, especially in the context of theories of truth.

Since we are working in a theory extending PA, we have all the formulas available that represent the primitive recursive syntactical relations for the language of arithmetic, especially:

$$(x, y) = z;\ Seq(x);\ lh(x) = y;\ Ct(x);\ Snt(x),\ Fml^1(x).$$

defining ordered pairs, sequences, the length of a sequence, closed terms, sentences and formulas with one free variable of the (truth-free) arithmetical language, respectively. We will use the metavariables s, t for closed terms, i.e. we will use $\forall t\,\psi(t)$ as an abbreviation for $\forall x(Ct(x) \rightarrow \psi(x))$. We use $\underline{n}$ to denote the numeral $S...S\underline{0}$, i.e. n successor symbols concatenated with the symbol for zero.

Furthermore we consider the language to contain function symbols for some specific primitive recursive functions. For some cases we use the dot notation $\dot{}$ such as $\dot{=}$; $\dot{\neg}$; $\dot{\wedge}$; $\dot{\vee}$; $\dot{\forall}$; $\dot{\exists}$; $\dot{ucl}$ for the representation of the functions that applied to the Gödel numbers of two terms s, t gives the Gödel number of the formula $s = t$, the negation, conjunction, disjunction, universal and existential quantification function and the function that takes the Gödel number of a formula to the Gödel number of its universal closure. We use the dot notation $\dot{x}$ for the representation of the *num*-function that takes a number to the Gödel number of its numeral. *val* represents the value function that applied to a closed term gives the value. Moreover we have a three place substitution function $x(t/s)$ that applied to a formula x a term t and a variable s gives the Gödel number of the formula resulting from substituting all free occurrences of s in x by t. In the case where x is a formula with one free variable, then $x(t)$ is short for the result of substituting the term t in x for all free occurrences of the free variable.[23]

The language $\mathcal{L}_T$ is the language of arithmetic $\mathcal{L}_A$ expanded by a one place predicate T. We intend T to be a truth predicate for the language $\mathcal{L}_A$.[24] The axioms of PT^- are given by the axioms of PA and the following axioms:

(*C1*) $\forall s\forall t(T(s \dot{=} t) \leftrightarrow val(s) = val(t))$

(*C2*) $\forall s\forall t(T(\dot{\neg}(s \dot{=} t)) \leftrightarrow \neg val(s) = val(t))$

(*C3*) $\forall x\forall y(Snt(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y)))$

(*C4*) $\forall x\forall y(Snt(\dot{\neg}(x \dot{\wedge} y)) \rightarrow (T(\dot{\neg}(x \dot{\wedge} y)) \leftrightarrow T(\dot{\neg} x) \vee T(\dot{\neg} y)))$

(*C5*) $\forall x(Snt(\dot{\forall}\, y\, x) \rightarrow (T(\dot{\forall}\, y\, x) \leftrightarrow \forall z T(x(\dot{z}))))$

(*C6*) $\forall x(Snt(\dot{\neg}\, \dot{\forall}\, y\, x) \rightarrow (T(\dot{\neg}\, \dot{\forall}\, y\, x) \leftrightarrow \exists z T(\dot{\neg} x(\dot{z}))))$

(*C7*) $\forall x(Snt(\dot{\neg}\, \dot{\neg} x) \rightarrow (T(\dot{\neg}\, \dot{\neg} x) \leftrightarrow T(x)))$

(*C8*) $\forall x(T(x) \rightarrow Snt(x))$

(*C9*) $\forall s\forall t\forall x(val(s) = val(t) \rightarrow (T(x(s)) \leftrightarrow T(x(t))))$

$$tot(x) :\Leftrightarrow Fml^1(x) \wedge \forall y(T(x(\dot{y})) \vee T(\dot{\neg} x(\dot{y}))).$$

[23] For precise definitions of these notions see Halbach (2014).

[24] This restriction is only for simplification and we could also extend the theory by giving axioms for languages containing $\mathcal{L}_A$.

The induction axiom is then a form of internal induction restricted to total formulas:

$$(I_tI) \qquad \forall x(tot(x) \wedge T(x(\underline{0})) \wedge \forall y(T(x(\dot{y})) \rightarrow T(x(y \dot{+} \underline{1}))) \rightarrow \forall y T(x(\dot{y})))$$

$$\mathrm{PT}^- := \mathrm{PA} \cup (C1) - (C9) \cup (I_tI)$$

The theory PT^- is adequate in Tarski's sense, i.e. all Tarski-biconditionals for $\mathcal{L}_A$ are provable. Moreover, all universal Tarski-biconditionals (sentences of the form $\forall x(\phi(x) \leftrightarrow T(\phi(\dot{x}))$, where ϕ does not contain T) are also provable.

PT^- is a compositional theory of truth. But in contrast to Tarskian theories of truth there is no axiom for the commutation of negation and truth. This commutation axiom is also called the 'completeness and consistency' axiom and is here abbreviated as (CC):

$$(CC) \qquad \forall x(Snt(x) \rightarrow (\neg T(x) \leftrightarrow T(\dot{\neg} x))).$$

The absence of (CC) means that PT^- is a *positive* theory of truth. The constructive part of the axioms is stated in a form such that the axioms can be rewritten as a positive inductive definition. Semantically, this gives rise to a monotone operator and a fixed point that defines the inductive set of true (in the standard model) arithmetical sentences.[25]

In contrast to its base theory, PT^- is finitely axiomatisable. For this it suffices to realise that $I\Sigma_1 \cup (C1) - (C9) \cup (I_tI)$ proves all the instances of arithmetical induction. As $I\Sigma_1$ is finitely axiomatisable, so is PT^-. The finite axiomatisability of PT^- is connected to the formulation of induction as an axiom instead of as a scheme. Induction formulated as an *axiom* is very natural: the disposition to accept all instances of a scheme seems to be rooted in the belief in a single principle. The induction axiom of PT^- is basically an induction principle for atomic formulas of the form $T(x)$. The point is that with our theory of syntax we can code sentences φ of arbitrary complexity. This complexity does not show up directly in the code $\ulcorner\varphi\urcorner$ so that $T(\ulcorner\varphi\urcorner)$ is still atomic. It is therefore natural that we get full arithmetical induction back even if we start only with induction for Σ_1-formulas. This also shows that, for any n, a restriction of induction to Σ_n-formulas is artificial as we will get full arithmetical induction by adding truth to it. So the choice of PA as base theory is justified and not arbitrary.

Replacing the induction schema by an induction axiom seems justified and there is a story to be told here which we will consider in the next section. But the induction axiom of PT^- is unusual: it is restricted to *total* formulas. At first sight this restriction might seem unjustified. In the following we will give three possible justifications.

Totality in itself seems to be a natural restriction at least in the context of type-free theories where we have cases of sentences that are neither true nor their negation is true. So Cantini and Feferman make use of such restrictions in their theories KF_t and DT.[26] The problem is that in the case of typed theories we seem to lack this motivation, since for all formulas $\phi(x)$ in the language of arithmetic and for all natural numbers n, $\phi(\underline{n})$ will be true in the standard model or its negation will be.

Nonetheless, this argument is not conclusive. An argument for the restriction of the induction principle goes as follows. We are considering a restriction of internal induction. In this case we have a formula $Fml^1(x)$ representing the set of arithmetical formulas with one free variable. By overspill we know that there are nonstandard models with

[25] The set of arithmetical truths is hyperelementary, because there is also a positive inductive definition that characterises the complement. See, for example, McGee (1991).

[26] In the setting of those type-free systems, adding (CC) leads to an outright inconsistency.

nonstandard elements satisfying this formula. Moreover it is possible to show that there are models $\mathfrak{M}$ of the positive inductive definition of truth with an element c such that $\mathfrak{M} \models Fml^1(x) \wedge \neg tot(x)[c]$. If we now understand $\{b \in M \mid \mathfrak{M} \models T(x(y))[c, b]\}$ as the extension and $\{b \in M \mid \mathfrak{M} \models T(\dot{\neg} x(y))[c, b]\}$ as the anti-extension of a nonstandard formula c in $\mathfrak{M}$, then this formula is flawed, i.e. there are b such that b is neither in the extension nor the anti-extension. The restriction of internal induction to total formulas is therefore a motivated restriction in the sense that it allows for induction only on formulas that are also well-behaved on the nonstandard part.

Thus, ultimately the situation is as follows. We do not want to accept instantiations of induction for non-standard elements that are not truth-determinate for the property in question for exactly the same reason that we resist inductive premises for soritical predicates. Note that the reply that in the "intended" model there are no such non-standard elements to be found is not undermining our motivation for restriction; as we have emphasised repeatedly, we are adopting the model-theoretic viewpoint, and from this viewpoint, all models are on a par.

A comparison with subsystems of second-order-arithmetic clarifies our point. In the case of the second-order theory of arithmetic ACA we have induction for subsets we do not have comprehension for: this reveals an unnatural asymmetry. This asymmetry is avoided in the case of the second-order arithmetical theory ACA_0.[27] So it is with PT^-. We only have induction for formulas that behave appropriately with respect to the truth predicate. The second-order theory ACA_0 is a sub-theory of a definitional extension of PT^-, and PT^- is a sub-theory of a definitional extension of ACA_0.[28] This is a much stronger connection than mutual relative interpretability. For example, the arithmetic vocabulary is not changed in the translation: the translation function induces a model transformation that preserves the interpretation of the arithmetical part.

PT^- is a semantically conservative extension of PA.[29] This fact is established indirectly as follows: Take a model $\mathfrak{M}$ of PA. As ACA_0 is a semantically conservative extension of PA we get a model $\mathfrak{M}^*$ of ACA_0.[30] By the fact that PT^- is a subtheory of a definitional extension of ACA_0 we have a transformation of ACA_0-models to PT^--models in such a way that the arithmetical structure is preserved. So we obtain a model $\mathfrak{M}'$ of PT^- and $\mathfrak{M}'$ is an expansion of $\mathfrak{M}$.

In a specific sense, PT^- is closely connected to its base theory: it does not restrict the possible interpretations and is therefore not substantial and in one sense only a minimal expansion of PA. In another respect it is not very close to PA, namely it is not relatively interpretable in PA. This is readily established by the finite axiomatizability of PT^- and Gödel's second incompleteness theorem. The fact that PT^- is not relatively interpretable in PA shows that PT^- is not conceptually reducible to it.[31] We explore the relation between this conceptual irreducibility and the gain in expressive power. The restriction of induction to total formulas is thus exactly what is needed to maintain the balance between semantical

---

27 See Simpson (1999).

28 For a proof of those two facts see Fischer (2009).

29 Cantini (1989) proves that the type-free theory KF_t containing PT^- is a syntactically conservative extension of PA. Although in one version of the conservativity proof he uses model-theoretic methods it does not directly imply the semantical conservativeness of PT^-.

30 See Simpson (1999).

31 For the connection between conceptual reducibility and relative interpretability, see Niebergall (2000).

conservativeness and non-interpretability in PA. So from a standpoint of metamathematical pragmatism, one can take the truth theoretic part as being purely instrumental over the arithmetical part. If one is prepared to take this standpoint, then the principles of the ideal part require no further motivation: success suffices.

PT^- is a typed theory of truth: it does not tell us anything about the truth of sentences that themselves contain the truth predicate. For explicating the implicit and explicit uses of the concept of truth in mathematical contexts, this is fine: in mathematics (in contrast to philosophy) there appear to be no uses of reflexive truth. Nonetheless, it is possible to extend PT^- to a natural type-free theory that captures the function of truth in mathematics equally well. The theory in question is called KF_t, which was first proposed and investigated by Cantini.[32] However, we will show in the following sections that the theory PT^- already fulfils the tasks that, from a general model theoretic perspective, one may expect a truth theory to be equal to. So for our purposes there seems no incentive to consider type-free variants of PT^- in further detail. Moreover, as far as the authors know, it is an open question whether KF_t is semantically conservative over PA.

§5. Finite axiomatisability. Kreisel held the view that our evidence for the truth of the first-order axiom scheme of mathematical induction resides in the second-order *axiom* of mathematical induction: we accept the first-order scheme *because* we believe the second-order axiom (as a single proposition).[33] But this has been disputed. We may not have a good grasp of the concept of being an arbitrary subset of an infinite set: this concept might be to some extent inherently indeterminate.

Nevertheless, Kreisel's theory brings a question about the relation between dispositions and beliefs to the fore, which is in the end a question about finitisation. An alternative to Kreisel's view would be the following. Our acceptance of the induction scheme for first-order arithmetical formulas is indeed backed up by a belief in a single proposition. But this proposition is merely what is expressed by the statement that all first-order arithmetical instances of the scheme are *true*.

A typical example of the expressive power of the truth predicate is given in examples such as 'All axioms of ZFC are true'. We can express in one sentence a statement that is not expressible (in a certain sense) in the language of set theory, which is usually considered to be very expressive.[34] ZFC itself is not finitely axiomatisable in the language of set theory and there is no finitely axiomatisable theory containing ZFC in the language of set theory. But with the truth predicate T and a definable predicate Ax_{ZFC} we can state a single sentence $\forall x\,(Snt\,(x) \wedge Ax_{\mathrm{ZFC}}(x) \rightarrow T\,(x))$, which is a fairly good approximation. To justify that it adequately approximates what we want, we need a theory capable of coding syntax and a theory of truth that satisfies convention T, such that all the uniform T-biconditionals are derivable. Furthermore the two theories should be finitely axiomatisable.

This is the problem about finite axiomatisability that Craig & Vaught (1958) sought to solve.[35] For any axiomatisable theory S with only infinite models, there is a finitely

---

[32] See Cantini (1989).

[33] See Kreisel (1967, p. 148).

[34] The sense in which it is not expressible in ZFC is the following: There is no sentence in the language of set theory having all axioms of ZFC as consequences, where the set of consequences is consistent.

[35] Their solution was based on a result by Kleene, and they actually proved a slightly more general theorem about finite axiomatisability than the one that we are discussing here.

axiomatisable theory S' in a language expansion, such that S' is a semantically conservative extension of S. The point is that the expansion is basically an expansion by syntactical predicates for $\mathcal{L}_S$ (the first-order language of set theory) and a predicate T, which is a truth predicate. The new axioms are the axioms of Q plus the compositional clauses for truth, which is sufficient to prove the universal T-sentences for $\mathcal{L}_S$, plus the statement $\forall x(Snt_{\mathcal{L}_S}(x) \wedge Ax_S(x) \rightarrow T(x))$. It is easily seen by the universal T-sentences that S' contains S and that it is finitely axiomatised. For the conservativeness proof Craig/Vaught use a model expansion argument.[36]

So the theorem given by Craig and Vaught is a very nice way to exemplify the expressive power of the truth predicate. Two questions arise: Are all theories of truth adequate for this purpose? What happens if the object theory is already capable of coding its own syntax?

Let us take the questions in turn. Concerning the first question, we have seen that for the purpose of a finite axiomatisation it is sufficient to have a finitely axiomatised theory of truth proving all the uniform T-sentences. This excludes theories of truth that have infinitely many truth axioms such as the disquotational theories TB and UTB. TB is the theory of truth extending PA by all the T-sentences for the language of arithmetic and UTB by all the uniform T-sentences. Both theories have full induction for the language $\mathcal{L}_T$.[37] In the case of compositional theories we have to make sure that the base theory is finitely axiomatisable. CT$\restriction$, i.e. the Tarskian compositional theory of truth with restricted induction, for example is not finitely axiomatisable, because PA is not. But the theory that Craig and Vaught use, which is $T(Q)\restriction$,[38] is finitely axiomatised. So in the case of Tarski's theory of truth, the finite axiomatisability depends on the choice of the base theory. This seems to exclude PA as a base theory for this purpose, although PA seems to be a very natural choice.

Concerning the second question, if we consider the exclusion of PA, then the remaining choices seem rather inadequate for the general case. For consider PA as object theory. Then PA itself is already capable of talking about its own syntax. So it is a viable option to take the object theory itself as base theory for the truth theory. But in this case an extension of PA by the axioms suggested by Craig/Vaught results in CT$\restriction$, which is not finitely axiomatisable and therefore cannot fulfil the expected purpose. The main reason is the infinite axiom schema of induction in PA. For a finite axiomatisation we would have to replace it by an axiom of induction as for example in systems of second-order arithmetic. But also the truth predicate allows the formulation of induction axioms.

So what we want for a non-finitely axiomatisable theory S containing Q, is a language expansion $\mathcal{L}_{S'}$ and a finitely axiomatisable theory S', such that S' is a (semantically) conservative extension of S, and S' makes use of the resources already given by Q. Specific examples of this phenomenon are PA and ACA_0 as well as ZFC and NBG. Note that if S is (essentially) reflexive, as is the case with PA and ZFC, then S does not interpret S' by Gödel's second incompleteness theorem. In both cases we find a conservative finite axiomatisation by increase of expressiveness. In both cases this expanded expressiveness seems to be gained by allowing second order variables. But what the second-order

---

[36] The compositional axioms are basically the axioms of CT$\restriction$ which correspond to the claim of the existence of a satisfaction class. Whereas CT$\restriction$ is not semantically conservative over PA because of Lachlan's theorem, Craig and Vaught's model expansion argument is compatible with Lachlan's result. The reason for this is that they introduce a new arithmetic theory and add therefore new syntactical vocabulary.

[37] For a precise definition see Halbach (2014).

[38] $T(Q)\restriction$ consists of Robinson Arithmetic (Q) plus the axioms that state that the truth predicate commutes with the logical connectives.

variables actually do is allowing a definition of a truth predicate which increases the expressive power. So this phenomenon is better understood as the possibility of defining a truth predicate. In both cases we can define a truth predicate for which the uniform T-biconditionals hold for the T-free instances. Moreover they satisfy some compositional clauses except for the commutation of truth with the negation sign.[39] So what we are looking for is a positive truth predicate.

We shall show that the truth theory PT^- is exactly of the desired form. PT^- is a finitely axiomatisable and a (semantically) conservative extension of PA, containing all universal Tarski-biconditionals. By applying these facts to our explication of expressivity we see that the truth predicate increases the expressive power. As classes of intended interpretations we consider for the language of arithmetic the class of Q models and for the language of truth the class of $\mathrm{TB}(Q)$ models, where $\mathrm{TB}(Q)$ is the disquotation theory of Tarski-biconditionals over the base theory Q.

The finite axiomatisation of PT^- shows that the class of PT^- models is elementary with respect to $\mathrm{TB}(Q)$, i.e. $\kappa = \{(\mathfrak{M}, S) \mid (\mathfrak{M}, S) \models \mathrm{PT}^-\}$ is EC in $\mathcal{L}_T$ with respect to $\mathrm{TB}(Q)$. $(\mathfrak{M}, S)$ is an expansion of the arithmetical model $\mathfrak{M}$ by a set S interpreting the predicate T. The set $\kappa' = \{(\mathfrak{M}, S)\restriction\mathcal{L}_A \mid (\mathfrak{M}, S) \in \kappa\}$ is by the semantical conservativity of PT^- over PA nothing but $\{\mathfrak{M} \mid \mathfrak{M} \models \mathrm{PA}\}$, which is not elementary in $\mathcal{L}_A$ with respect to Q, since PA is not finitely axiomatisable. This argument establishes then the following observation.

PROPOSITION 5.1. *The language of truth (with respect to* $\mathrm{TB}(Q)$*) is expressively stronger than the language of arithmetic (with respect to* Q*), i.e.*

$$\mathcal{L}_A(Q) < \mathcal{L}_T(\mathrm{TB}(Q)).$$

A comment on the role of TB is in order to guard against misunderstandings of this proposition. At first glance the axioms of TB might appear to be superfluous and the proposition could be trivialized by the fact that we can establish a similar proposition for $\mathcal{L}_T$ with respect to Q formulated in $\mathcal{L}_T$. But in fact only the restriction via TB axioms gives the proposition the specific content we want, namely that it is a truth predicate which increases the expressive power. Without any specific axioms characterizing the new predicate T it only says that there is some predicate that allows for more expressive power. The truth predicate is so to say a witness for this phenomenon.

Similarly it is the role that PT^- plays in the proof that highlights its merits, namely as a witness for a finitely axiomatizable theory conservatively extending PA. Other conservative theories of truth like UTB and $\mathrm{CT}\restriction$ are not able to capture this aspect of the expressive function of truth as they are neither finitely axiomatisable nor semantically conservative.

There are other contexts in which something similar might prove useful. If we compare our case with the case of set theory one could argue that our justification for accepting the schema of replacement lies in the acceptance of a second-order axiom. So from a Kreiselian perspective it is the acceptance of a finitely axiomatised theory like NBG that justifies us in our acceptance of ZFC. But in the case of NBG this would entail an ontological commitment to proper classes. To avoid this ontological commitment one could try to replace the class theory by a theory of satisfaction that is a finitely axiomatisable conservative extension of ZFC. How this is done is shown by Fujimoto (2012, chap. 6.2.), where he shows that the natural counterpart to Cantini's KF_t for ZFC, which Fujimoto calls KF_{tc}, has similar properties. Especially the conservativity and finite axiomatisability

[39] For a summary, see Schindler (1994).

can be established. It is easy to see that the natural typed counterpart for ZFC of the typed theory PT^- for ZFC has the same properties.

§6. Proving semantical metatheorems. In this section we show how to make use of the expressive function of the truth predicate as captured by PT^- to prove important semantical metatheorems. This desideratum is already given in Ketland (1999), and it is one of the key tasks that we expect a truth theory to be able to carry out. We focus on Gödel's completeness theorem. For a given logical calculus, the completeness is usually stated as saying that: If $\Sigma \models \varphi$, then $\Sigma \vdash \varphi$, or, equivalently, every consistent set of sentences Σ has a model (making all the sentences in Σ true).[40]

Nowadays a very common way to proof completeness follows Henkin's proof by building term models. Usually the textbook proofs are done informally, but it is tacitly assumed that they can be formalised as proofs in ZFC. Furthermore it is well known that ZFC is much stronger than is necessary for this task: already subsystems of second-order arithmetic are sufficient to carry out the proof.[41] But already in first-order arithmetical systems weaker versions of the completeness proof can be carried out.

So let us review the completeness proof using Henkin models. Take a set of $\mathcal{L}$-sentences Σ that is consistent. Σ can be extended to a maximally consistent set of sentences Σ' in a language expansion $\mathcal{L}'$ of $\mathcal{L}$ by a set of new constants. Furthermore the maximally consistent set Σ' can be so construed as to contain witnesses for all existence claims. This allows one to build a model $\mathfrak{M}$, where the domain is the set of terms of $\mathcal{L}'$ and a sentence is true in $\mathfrak{M}$ iff it is in Σ'. So we have constructed a Henkin model satisfying Σ. This proof is formalisable in set theory. We divide the completeness proof into three steps: Step 1. Constructing a conservative Henkin expansion; Step 2. Every consistent set can be extended to a maximally consistent set; Step 3. Every Henkin set has a term model.

In the following we will sketch a proof following a formalisation in second order arithmetic: for a more detailed proof see Section 12. First of all we will use the notion of truth to recover the relevant second order notions. For this we will translate the language of second-order arithmetic into the language $\mathcal{L}_T$. It is then possible to show that we can recover arithmetical comprehension in PT^- as a form of open truth comprehension. This form of comprehension is sufficient to proof a translation of König's lemma in PT^-. König's lemma is relevant because a variant of the completeness proof that makes use of trees is more fitting for a formalisation in theories weaker than set theory, such as subsystems of second-order arithmetic.

The notions recovered by the translation are sufficient to carry out the first step of the completeness proof in PT^-. However the crucial step in the completeness proof is the second step which is handled by an application of Lindenbaum's lemma. The basic idea in the formalisation is to reduce the question to trees, because then Lindenbaum's lemma can be reconstructed via König's lemma.

In our version we can define the notion of a truth set, which is basically a consistent set closed under the positive clauses of truth. It is a full truth set if it also contains all the true

---

[40] Since Gödel first proved the completeness of first-order logic in his dissertation in 1929, there have been further conceptual developments in this area. One of them is Tarski's definition of a model and the notion of truth in a model.

[41] Simpson (1999, chap. IV.3) shows that WKL_0, a system based on weak König's lemma, is sufficient.

literals. We can then prove that every consistent set can be expanded to a full truth set. This is established as follows: For every set of sentences Σ we can construct a tree T_Σ in such a way that T_Σ is infinite iff Σ is consistent. Moreover the tree is such that an infinite path through the tree will be a full truth set containing Σ. Now if Σ is consistent, then T_Σ will be infinite. By König's lemma we can find an infinite path through that tree which will be a full truth set.

These full truth sets are of such a form that they can be easily transformed into models, which establishes the third step. This concludes the completeness proof in PT^- and shows that the truth predicate functions here in a natural way in the proof of a generalized version. This is an example of the expressive function captured by PT^-.

Let us now compare our formalisation in our preferred theory of truth to other versions. Already in axiomatic first-order arithmetics we can prove restricted versions of the completeness theorem. Because of the limited expressive power, we can prove completeness only for definable sets of sentences. In PA it is possible to prove a completeness theorem for arithmetically definable theories.[42] A version of the arithmetized completeness theorem can already be proved in $I\Sigma_1$ for recursively axiomatised theories.[43] These first-order arithmetics are expressively not sufficient to prove the generalised versions of the theorem; we can only prove a schematic form. More promising as adequate formalisations are versions in second order arithmetic.

Already RCA_0 is sufficient for the definition of the necessary concepts and to prove some of the relevant properties. In RCA_0 we can define the completion of a set of sentences by witnesses for existential statements to execute the first step and the third step in the completeness proof. But we can only prove the existence of a completion of Σ for a consistent and deductively closed Σ. This gives us only a weak version of the completeness theorem in RCA_0: it gives us a completeness theorem only for deductively closed sets of sentences.[44] Recursive comprehension is not enough for the general case as there are recursive sets of sentences that do not have a recursive completion, such as Q. For the proof of the existence of a maximal consistent set with the right properties, weak König's lemma is both sufficient and necessary.

Weak König's lemma allows us get from an infinite binary tree T_Σ to an infinite path in T_Σ. This infinite path gives rise to a model satisfying Σ. So in WKL_0 we have a generalised version of the completeness theorem for arithmetically definable sets of sentences.[45] WKL_0 is a fairly weak system of second-order arithmetic. The first-order part of WKL_0 is the same as the first-order part of RCA_0, namely $I\Sigma_1$. Furthermore, WKL_0 is interpretable in PA.[46] So PA can simulate this completeness proof.

In the usual proof for countable languages an explicit construction of a maximally consistent set is given, although the existence of such a set is sufficient. This difference becomes important when formalising the completeness proof.

Trivially we can prove the same completeness theorem in stronger systems of second-order arithmetic, especially in ACA_0. But there is an advantage of the proof in ACA_0. ACA_0 is equivalent to (full) König's lemma over RCA_0. Whereas with weak König's

---

42 Kaye (1991, p. 186).

43 (Hájek & Pudlák, 1993, p. 104). It is interesting to notice that in this proof the low basis theorem is used, which is in close relation to Weak König's lemma.

44 Simpson (1999, p. 92).

45 Simpson (1999, p. 139).

46 See Hájek (1993, p. 189f).

lemma only the existence of an infinite path can be proved without giving a way to pick out this infinite path, we can *construct* the infinite path in ACA_0 in the sense that we can pick out one unique path, for example the leftmost.[47] In most of the textbook presentations the Lindenbaum lemma is proved in such a constructive way that will provide a maximally consistent set. In such a case the formalisation in ACA_0 seems more faithful. It is known that PA does not interpret ACA_0. In this sense, even in the usual completeness proof concepts are used that exceed those of PA. But note that all these theories are arithmetically conservative over PA.

ACA_0 is a natural fragment of second-order arithmetic (a natural 'stopping point'). It plays an important role in reverse mathematics. It is well known that there are many theorems that are provably (in a weak theory) equivalent to ACA_0 and some of them are also relevant for semantic notions such as truth. Examples are versions of Ramsey's theorem and the Ehrenfeucht–Mostowski lemma.[48] It is an advantage of a theory of truth to be able to recover all these theorems. Theories of truth that are interpretable in the base theory PA cannot capture these theorems, as the following argument shows.

Assume that T is a theory of truth interpretable in PA and A is a theorem that is equivalent to ACA_0 i.e. $\mathrm{RCA}_0 \vdash A \leftrightarrow \mathrm{ACA}_0$. Assume further that there is a suitable interpretation I of the theorem A in T. This means that that there is a structure preserving translation from $\mathcal{L}_A^2$ in $\mathcal{L}_T$, such that $T \vdash I(A \wedge \mathrm{RCA}_0)$ and for all B, if $\mathrm{RCA}_0, A \vdash B$, then $T \vdash I(B)$. But the equivalence implies that $\mathrm{RCA}_0, A \vdash \mathrm{ACA}_0$ and therefore $T \vdash I(\mathrm{ACA}_0)$, which would mean that T interprets ACA_0. But as T is interpretable in PA and interpretability is transitive ACA_0 would also be interpretable in PA, which is not the case.

The proof of the completeness theorem in PT^- exemplifies the role a truth predicate is able to play. PT^- is not a theory capable of formalizing all model theoretic results and it is not argued for a replacement of standard model theory formalized set theoretically. Rather the purpose of the outlined proof is to show that basic results of model theory, which rely on the notion of truth, can be carried out in a minimal theory of truth. This is again evidence that the concept of truth of PT^- has genuine expressive power that is not captured by PA or by theories of truth that are interpretable in PA, such as $\mathrm{CT}\restriction$.

§7. Expressiveness exemplified by speed-up. The expressive function of the truth predicate is also explicable from a pragmatic point of view. We can think of the function as being merely instrumental: whereas the truth predicate does not allow to prove new arithmetical theorems it allows us to simplify proofs of arithmetical theorems already derivable in PA. The expressive strength of the truth predicate lies in its ability to shorten proofs.

This pragmatic understanding of expressiveness connects deflationary conceptions of truth with tennets of instrumentalism especially with Hilbert's program. Conservativity plays also in Hilbert's program a justificatory role for the ideal part of mathematics. The instrumental function of the ideal part is sometimes connected with its role in simplifying proof procedures of theorems of the real part.[49] Something analogous is the case with minimal theories of truth. The truth predicate is able to simplify proofs of the base theory

[47] In order to avoid confusion we have to say that this use of construct is not exactly the one of constructivist mathematics, but rather a more naive use of construct.

[48] Simpson (1999, p. 122).

[49] Compare Caldon & Ignjatovič (2005).

without committing itself to new theorems in the base language. This instrumental function of the truth predicate is part of its expressive function.

In our case the instrumental function of simplifying proofs of the truth predicate can be established as the following claim: the theory PT^- has non-elementary speed-up over PA, i.e., there is an arithmetical formula $\varphi(x)$, such that $\mathrm{PA} \vdash \varphi(\underline{n})$ for all $n \in \omega$, but the length of proofs for $\varphi(\underline{n})$ in PA for these statements are not bounded by any function f polynomial in n, but only superexponential in n. On the other hand we find relatively short proofs for those statements in PT^-. This is exactly one of the advantages a purely expressive device could have: we cannot prove new theorems in the old language, but for the theorems we can prove we find shorter and more natural proofs. There are well known unpublished speed-up results by Solovay for ACA_0 over PA as well as for BG over ZF and the following speed-up proof relies on the method developed by Solovay and explained in Pudlák (1998).

THEOREM 7.1. PT^- *has non-elementary speed-up over* PA.

In the following we will sketch the main ideas of the proof.[50] The formula we want to show speed-up for is a restricted consistency statement $Con_{pa}(x)$ saying that there is no proof of length less or equal to x of $0 = 1$. Friedman and Pudlák showed that a proof of $Con_{pa}(\underline{n})$ in PA contains at least n symbols. So if we plug in the super-exponential function, then a proof of $Con_{pa}(2_{\underline{n}})$ in PA contains at least $2_{\underline{n}}$ symbols.

On the other hand we can give polynomial bounds in n on the proofs in PT^- of $Con_{pa}(2_{\underline{n}})$. This is due to the fact that we can define a cut $C(x)$ in PT^- with certain properties. A cut is basically an initial segment of a model of arithmetic that satisfies certain closure conditions, such as that if a number (standard or non-standard) is in the cut, then so is its successor and if a number is in the cut then all numbers smaller are also in the cut. By using the method of shortening cuts that was developed by Solovay, for such a cut $C(x)$ we can prove $C(2_{\underline{n}})$ in PT^- by a proof of size polynomial in n. The crucial step is then the following lemma:

LEMMA 7.2. *There is a definable cut* $C(x)$ *in* PT^- *such that* PT^- *proves the consistency of* PA *on this cut, i.e.*

$$\forall x(C(x) \rightarrow Con_{pa}(x)).$$

Proof. (Sketch) The basic idea is to define a cut such that all formulas in this cut are total. We can use this cut in order to prove a restricted reflection principle for this cut. That is, we can prove that all axioms of PA are true on this cut and that the rules of inference preserve truth on this cut. This is sufficient to define a new cut, such that all PA-provable sentences on this cut are true. So we can prove that PA is consistent on this cut. □

In contrast to the speed-up of PT^- over PA, the disquotational theories like TB and UTB have no meaningful speed-up. For those theories we have a simple way to locally interpret them in PA and keeping the arithmetical vocabulary fixed, namely by replacing, in every proof, the axiomatic truth predicate by the partial truth predicate that is definable in PA and satisfies all the Tarski-biconditionals that are used in the proof. To transform it into a proof in PA we have to add the proofs for the Tarski-biconditionals of the defined predicates. But these proofs can be bound by a polynomial, and therefore this also holds for the length of whole proof.

[50] The details of this proof are worked out in Fischer (2014).

Let us now compare PT^- in terms of speed-up with classical compositional truth. Generally the question whether $\mathrm{CT}\restriction$ has non-elementary speed-up over PA is open as far as is known to the authors. But there are some partial results: If we add the statement that all axioms are true, then we can prove the consistency on a cut and thereby obtain speed-up.[51] If we do not, then we can see that the interpretability of $\mathrm{CT}\restriction$ in PA excludes Π_1 sentences to be witnesses for the speed-up.[52] So if there is speed-up, then we would need a conceptually different proof.

§8. Reflection principles. The restriction to conservative extensions of PA has well-known limitations for a theory of truth. We want to address one of the main concerns one might have.[53] The truth predicate is a natural candidate for expressing the soundness of the base theory that we accept. The problem is that by Gödel's second incompleteness theorem a conservative extension of PA is not able to prove a standard consistency statement for PA. This also implies that a conservative theory of truth containing UTB cannot prove a soundness statement in form of the global reflection principle $\forall x(Snt(x) \wedge Pr_{pa}(x) \rightarrow T(x))$.

This way of expressing the soundness of arithmetic excludes models. In particular, it excludes those non-standard models that contain 'proofs' (of non-standard length) of $0 = 1$. From an algebraic perspective, this is not acceptable: each model of PA is as good as any other.

Viewed in this light, it is not at all clear that the global reflection principle is 'the' correct way to state soundness or the only one. PT^- cannot prove the global reflection principle in its general form. But if we restrict the formulas in a specific way as is done in Lemma 7.2, then PT^- is able to prove a restricted version of global reflection:

$$\forall x(Snt(x) \wedge \exists y(Proof_{pa}(y, x) \wedge C(x) \wedge C(lh(y))) \rightarrow T(x)).$$

This restricted version of global reflection is also adequate to state soundness. For one thing we know that all standard numbers are in the cut and therefore the proof predicate on a cut is extensionally correct in the sense that it strongly represents the relation of being a proof in PA. So we can prove that there is no proof in PA of an inconsistency that has standard length. In sum, even if one thinks that a reasonable theory of truth is committed to prove a reflection principle for the base theory to be able to express a soundness claim, one can still argue (even though it is a conservative extension of PA) that PT^- is adequate in this respect.[54]

§9. Conclusion. There are two perspectives from which a mathematical theory such as arithmetic or set theory can be viewed. On the one hand, one can take these theories to be about one privileged model or about a small class of privileged models. On the other

---

51 This unpublished result was reported by Albert Visser.

52 This observation is also due to Albert Visser.

53 These concerns were raised a.o. by Ketland (1999) and Shapiro (1998).

54 Whether provability on a cut is an intensionally correct representation of provability *simpliciter* is debatable. Sometimes the Löb conditions are taken to be criteria of adequacy, also because provability predicates satisfying them are subject to Gödel's second incompleteness theorem. Provability on a cut satisfies at least Löb 1. Moreover it satisfies weaker versions of Löb 2 and Löb 3 making it subject to a strengthened version of Gödel's second incompleteness theorem. See Hájek & Pudlák (1993, chap. 3(b)).

hand, one can take all models of these theories to be on a par. The latter can be called the model-theoretic or algebraic perspective.

It is the latter viewpoint that we have adopted in this article. We have not defended this viewpoint here; we have contented ourselves with observing that this is a viewpoint that is commonly adopted in certain parts of mathematics such as model theory.

Now suppose we want to formulate an adequate truth theory for a mathematical theory. If the algebraic perspective is adopted, then the truth theory ought to be semantically conservative. When added to the theory, no models of the original theory ought to be thereby ruled out, otherwise it is not a truth predicate *for* that theory. At the same time, it is generally recognised that the central role of the concept of truth is to fulfil an expressive role: it ought to allow us to express certain propositions (mostly semantical propositions) that we could not express before.

The central claim of this article is that there are natural truth theories that satisfy both the demand of semantical conservativeness and the demand of adequately extending the expressive power of our language. In particular, the theory PT^- fils the bill. It is semantically conservative over its mathematical base theory. Its expressive power is most clearly witnessed by its conceptual irreducibility to (or non-interpretability in) the base theory.

The expressive power of the truth predicate also outs itself in a number of concrete phenomena. The theory PT^- allows the finite expression of the content of mathematical theories that are not finitely axiomatisable. PT^- accommodates natural proofs of semantic theorems (the completeness theorem), and it has non-elementary speed-up with respect to its arithmetical base theory. The technical results in this article are not particularly deep: they follow relatively straightforwardly from results that are well known. Instead, the aim of this article was to give a philosophical discussion of the minimal requirements that a truth theory has to satisfy for the purposes of doing metamathematics. To do this, we needed to bring out explicitly the role that truth plays in metamathematical reasoning.

PT^- is not the only truth theory that satisfies these two core requirements, and for some purposes it is perhaps not even the most attractive such theory. There might, for instance, be reasons for preferring some type-free extension of it. But PT^- does appear to be the *minimal* truth theory that satisfies these (from an algebraic perspective) core demands. It has the properties that one minimally requires from a truth theory when one takes a model-theoretic perspective, and it has no more than these properties.

§10. Acknowledgments. Martin Fischer's work was carried out within the project 'Syntactical treatments of interacting modalities' funded by the DFG. The project is hosted by the MCMP which is supported by the Alexander von Humboldt foundation. The authors want to thank the audiences of Bristol, Amsterdam, Leuven, Munich, Neuchatel, and Vienna where earlier versions of this paper were presented.

§11. Criterion of Expressiveness. Let $\mathrm{IM}_{\mathcal{L}} \subseteq \mathrm{MOD}^{\mathcal{L}}$. We say that a class of $\mathcal{L}$-models κ is EC (an Elementary Class) in $\mathcal{L}$ with respect to $\mathrm{IM}_{\mathcal{L}}$ if and only if there is a formula $\varphi \in \mathcal{L}$ such that $\kappa = \mathrm{MOD}^{\mathcal{L}}(\varphi)$ and $\mathrm{MOD}^{\mathcal{L}}(\varphi) \subseteq \mathrm{IM}_{\mathcal{L}}$. When we compare language expansions we have to make sure that we also increase the class of intended structures, i.e. if we have two languages $\mathcal{L}_1 \subseteq \mathcal{L}_2$ we have to assume that $\mathrm{IM}_{\mathcal{L}_1} \subseteq \{\mathfrak{M}\restriction\mathcal{L}_1 \mid \mathfrak{M} \in \mathrm{IM}_{\mathcal{L}_2}\}$. This allows us to formulate the following explication:

Explication. *Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$ be two first-order languages. Let $\mathrm{IM}_{\mathcal{L}_1} \subseteq \mathrm{MOD}^{\mathcal{L}_1}$ and $\mathrm{IM}_{\mathcal{L}_2} \subseteq \mathrm{MOD}^{\mathcal{L}_2}$, such that $\mathrm{IM}_{\mathcal{L}_1} \subseteq \{\mathfrak{M}\restriction\mathcal{L}_1 \mid \mathfrak{M} \in \mathrm{IM}_{\mathcal{L}_2}\}$. We say that $\mathcal{L}_2$ with respect to $\mathrm{IM}_{\mathcal{L}_2}$ is **expressively stronger** than $\mathcal{L}_1$ with respect to $\mathrm{IM}_{\mathcal{L}_1}$, in symbols*

$\mathcal{L}_1(\mathrm{IM}_{\mathcal{L}_1}) < \mathcal{L}_2(\mathrm{IM}_{\mathcal{L}_2})$, *iff there is an elementary class* $\kappa \subseteq \mathrm{IM}_{\mathcal{L}_2}$, *such that* $\kappa' = \{\mathfrak{M}\restriction\mathcal{L}_1 \mid \mathfrak{M} \in \kappa\}$ *is not elementary in* $\mathcal{L}_1$ *with respect to* $\mathrm{IM}_{\mathcal{L}_1}$.

§12. Completeness. For significant parts of this appendix we rely on definitions and proofs from Simpson (1999). We first show how to recover a fragment of second-order arithmetic in PT^-. We talk about sets by translating second-order arithmetic into the language of truth in the following way. $\sigma : \mathcal{L}_A^2 \to \mathcal{L}_T$:

$$\sigma(x_i) := x_{2i}; \sigma(X_j) := x_{2j+1};$$
$$\sigma(x_i \in X_j) := T(\sigma(X_j)(\sigma(\dot{x}_i))) = T(x_{2j+1}(\dot{x_{2i}}))$$

To simplify things we will use greek letters α, β, γ as variables for total arithmetical formulas. So $\forall\alpha\varphi(\alpha)$ is an abbreviation for $\forall x(tot(x) \to \varphi(x))$. In this notation, our induction axiom can be written as:

$$(I_tI) \qquad \forall\alpha(T(\alpha(\underline{0})) \wedge \forall y(T(\alpha(\dot{y})) \to T(\alpha(y \dot{+} \underline{1})) \to \forall y T(\alpha(\dot{y}))).$$

This allows us to define the following notions: $\alpha \subseteq \beta$, α is finite, $k \in \alpha \times \beta$, f is a (total) function from α to β and f is an n-ary function $f : \mathbb{N}^k \to \mathbb{N}$.

A formula $\varphi \in \mathcal{L}_T$ is an *open truth formula* if and only if all the formula-variables α occurring in φ inside the truth predicate are free and total. Now we can prove some basic facts, such as open truth comprehension:[55]

THEOREM 12.1.

$$\exists\alpha\forall y(T(\alpha(\dot{y})) \leftrightarrow \varphi(y))$$

for all open truth formulas φ, *such that* α *not free in* φ.

We can prove some basic facts about functions in PT^-, for example that composition, primitive recursion and minimisation give new functions. Moreover we can define trees and paths in a way that we can prove König's lemma in PT^-.[56]

THEOREM 12.2 (König's lemma). *Every infinite finitely branching tree* $\mathcal{T}$ *has at least one infinite path.*

These are the basics which we need in the following. For the completeness theorem we want to consider various languages and theories formulated in those languages. The signature of a language $\mathcal{L}$ is a set of relation symbols, function symbols and constant symbols. The set of terms and the set of formulas of $\mathcal{L}$ are defined inductively as usual.

For a language $\mathcal{L}$ we assume a fixed signature and symbol numbering. In PT^- we can prove the existence of a Gödel numbering for all expressions of $\mathcal{L}$ by primitive recursion. We can then identify expressions with their Gödel numbers as usual and prove in PT^- the existence of formulas $Term_{\mathcal{L}}$, $Form_{\mathcal{L}}$ strongly representing the relations of being an $\mathcal{L}$-term and $\mathcal{L}$-formula by open truth comprehension taking $\mathcal{L}$ as a parameter. We also get provability predicates for arithmetically definable sets of formulas.

Let α be a formula defining a set of $\mathcal{L}$-formulas. Then we can express 'x is a proof of y from α' and 'y is derivable from α':

---

55 For a proof see Fischer (2009).

56 The proof is similar to the proof in ACA_0, as given in Simpson (1999, p. 121)

$$Proof_{\mathcal{L}}(\alpha, x, y) :\leftrightarrow Seq(x) \wedge \forall k < lh(x)(Fml_{\mathcal{L}}((x)_k) \wedge y = (x)_{lh(x)-1} \wedge (T(\alpha(\dot{(x)}_k)) \vee lAx((x)_k) \vee \exists i, j < k((x)_k = (x)_i \dot{\rightarrow} (x)_j))$$

$$Pr_{\mathcal{L}}(\alpha, y) :\leftrightarrow \exists x(Proof_{\mathcal{L}}(\alpha, x, y))$$

Clearly if $Proof_{\mathcal{L}}(\alpha, x, y)$, then $Proof_{\mathcal{L}'}(\alpha, x, y)$ for all $\mathcal{L} \subseteq \mathcal{L}'$.

DEFINITION 12.3. *A formula α defines a truth set for $\mathcal{L}$ iff*

$$\begin{aligned}
&\forall y(T\alpha(y) \rightarrow Snt_{\mathcal{L}}(y)) &\wedge\\
&\forall y \neg(T\alpha(y) \wedge T\alpha(\dot{\neg} y)) &\wedge\\
&\forall y\, (T\alpha(y) \leftrightarrow T\alpha(\dot{\neg}\, \dot{\neg} y)) &\wedge\\
&\forall y, z\, (T\alpha(y) \wedge T\alpha(z) \leftrightarrow T\alpha(y \dot{\wedge} z)) &\wedge\\
&\forall y, z\, (T\alpha(\dot{\neg}(y \dot{\wedge} z)) \leftrightarrow T\alpha(\dot{\neg} y)) \vee T\alpha(\dot{\neg} y))) &\wedge\\
&\forall y\, (T\alpha(\dot{\forall} zy) \leftrightarrow \forall z T\alpha(y(\dot{z}))) &\wedge\\
&\forall y\, (T\alpha(\dot{\neg}\, \dot{\forall} zy) \leftrightarrow \exists z T\alpha(\dot{\neg} y(\dot{z})))
\end{aligned}$$

A full truth set x is a truth set such that for all literals l either $l \in x$ or $\neg l \in x$.

Second part:

Let us now sketch a proof of the completeness in PT^-. Consider the first step: Let $\mathcal{L}$ be a first-order language. Take C to be a countably infinite set of constants not in $\mathcal{L}$. $\mathcal{L}_C = \mathcal{L} \cup C$. Let $\langle \varphi_n \mid n \in \mathbb{N}\}$ be an enumeration of formulas of $\mathcal{L}_C$ with one free variable and $\langle \underline{c_n} \mid n \in \mathbb{N}\}$ a one-to-one enumeration of C, such that for all n, $\underline{c_n}$ does not occur in any φ_i with $i \leq n$. The Henkin sentences μ_n are sentences of the following form: $\mu_n = \exists x \varphi_n(x) \rightarrow \varphi(\underline{c_n})$. Let Σ be a set of $\mathcal{L}$-formulas. The set Σ^*, Σ extended by all Henkin sentences, is called the Henkin extension of Σ in $\mathcal{L}_C$.

LEMMA 12.4. *(PT^-)*
If Σ is a consistent set of formulas, then its Henkin extension Σ^* is also consistent.

Proof. Let Σ be a consistent set of $\mathcal{L}$-formulas. One can easily prove in PT^- that Σ is also consistent in $\mathcal{L}_C$. By the formalization of the usual syntactical argument it follows that Σ^* is consistent. □

For the second step we prove a version of Lindenbaum's lemma in the following way.

LEMMA 12.5. *(PT^-) For every consistent set Σ, there is a full truth set $\Sigma^* \supseteq \Sigma$.*

Proof. Define a binary branching tree T_Σ of truth sets in the following way: $t \in T_\Sigma$ iff

$$\begin{aligned}
&\forall n < lh(t)((t)_n = 1 \rightarrow Snt_{\mathcal{L}}(n)) &\wedge\\
&\forall n < lh(t)(n \in \Sigma \rightarrow (t)_n = 1) &\wedge\\
&\forall n, m < lh(t)(n = \dot{\neg} m \rightarrow ((t)_n = 1 \rightarrow (t)_m = 0)) &\wedge\\
&\forall n, m < lh(t)(n = \dot{\neg}\, \dot{\neg} m \rightarrow ((t)_n = 1 \leftrightarrow (t)_m = 1)) &\wedge\\
&\forall k, n, m < lh(t)(k = m \dot{\wedge} n \rightarrow ((t)_k = 1 \leftrightarrow (t)_n = 1 \wedge (t)_m = 1)) &\wedge\\
&\forall k, n, m < lh(t)(k = m \dot{\vee} n \rightarrow ((t)_k = 1 \leftrightarrow (t)_n = 1 \vee (t)_m = 1)) &\wedge\\
&\forall k, n, m < lh(t)(k = \dot{\neg}(n \dot{\vee} m) \leftrightarrow ((t)_k = 1 \leftrightarrow (t)_n = 0 \wedge (t)_m = 0)) &\wedge\\
&\forall k, n, m < lh(t)(k = \dot{\neg}(n \dot{\wedge} m) \rightarrow ((t)_k = 1 \leftrightarrow (t)_n = 0 \vee (t)_m = 0)) &\wedge\\
&\forall n, m < lh(t)(n = \dot{\forall} zm \rightarrow ((t)_n = 1 \leftrightarrow \\
&\qquad \forall z(m(\dot{z}) < lh(t) \rightarrow (t)_{m(\dot{z})} = 1))) &\wedge
\end{aligned}$$

$$\forall n < lh(t)(n = \dot{\neg} \dot{\forall} zm \rightarrow ((t)_n = 1 \leftrightarrow \exists z(m(\dot{z}) < lh(t) \wedge (t)_{m(\dot{z})} = 0)))$$

The tree T_Σ exists by open truth comprehension. Moreover, it is provable in PT^- that Σ is consistent iff T_Σ is infinite.

Define T^* to be the set of nodes in T that are infinitely expandable, i.e. $t \in T^*$ iff $t \in T_\Sigma$ and the set of $\sigma \in T_\Sigma$, such that $t \subseteq \sigma$, is infinite. T^* exists by open truth comprehension. T^* is nonempty as T_Σ is infinite and therefore contains $\langle\rangle$. But in every step in T^* there are at most two immediate successors and at least one. So we can define a function $f(t) = t^\frown\langle m\rangle$, where m is the least number n, such that $t^\frown\langle n\rangle \in T^*$. By primitive recursion we can define a path g through T_Σ, $g : \mathbb{N} \rightarrow \mathbb{N}$, such that $g[n] = \langle g(0), ..., g(n-1)\rangle$. This infinite path g defines a full truth set. □

As for the third step:

DEFINITION 12.6. *A countable model $\mathfrak{M}$ of signature $\mathcal{L}$ is an ordered pair (M, I), where $M \subseteq \mathbb{N}$ and I is a function assigning to each constant symbol c an element of M, each n-place relation symbol a n-ary relation and to each m-ary function symbol a m-ary function. A variable assignment is a function $\sigma : Var \rightarrow M$. A satisfaction relation for $\mathfrak{M} \models \varphi[\sigma]$ has to satisfy the usual Tarski clauses:*

$$\mathfrak{M} \models s = t[\sigma] \Leftrightarrow s^{\mathfrak{M},\sigma} = t^{\mathfrak{M},\sigma}$$

LEMMA 12.7. *Every full truth set induces a model.*

Proof. Take $M := \{\underline{c_n} \mid \neg\exists m < n(n = m \in \Sigma^*)\}$ and $c^{\mathfrak{M}} :=$ the least $\underline{c_n}$ such that $c = c_n \in \Sigma^*$. For a n-ary relation symbol P take

$$P^{\mathfrak{M}} := \{\langle c_1, ..., c_n\rangle \mid c_1, ..., c_n \in M \text{ and } P(c_1, ..., c_n) \in \Sigma^*\}.$$

For f a m-ary relation symbol take

$$f^{\mathfrak{M}} = \{\langle c_1, ..., c_{m+1}\rangle \mid c_1, ..., c_{m+1} \in M \text{ and } f(c_1, ..., c_m) = c_{m+1} \in \Sigma^*\}.$$

$\mathfrak{M}$ exists by open truth comprehension and we can prove that the Tarski clauses hold. □

THEOREM 12.8. *(PT^-) Every consistent set of sentences is satisfiable.*

BIBLIOGRAPHY

Barwise, J., & Feferman, S. (1985). *Model-Theoretic Logics*. New York: Springer-Verlag.

Caldon, P., & Ignjatovič, A. (2005). On mathematical instrumentalism. *Journal of Symbolic Logic*, **70**(3), 778–794.

Cantini, A. (1989). Notes on formal theories of truth. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **35**, 97–130.

Craig, W., & Vaught, R. L. (1958). Finite axiomatisability using additional predicates. *The Journal of Symbolic Logic*, **23**, 289–308.

Fischer, M. (2009). Minimal truth and interpretability. *Review of Symbolic Logic*, **2**(4), 799–815.

Fischer, M. (2014). Truth and speed-up. *Review of Symbolic Logic*, **7**(2), 319–340.

Fujimoto, K. (2012). Classes and truths in set theory. *Annals of Pure and Applied Logic*, **163**, 1484–1523.

Hájek, P. (1993). Interpretability and fragments of arithmetic. In Clote, P. and Krajíček, J., editors. *Arithmetic, Proof Theory, and Computational Complexity*. Oxford: Oxford University Press, pp. 185–196.

Hájek, P., & Pudlák, P. (1993). *Metamathematics of First-Order Arithmetic*. Berlin: Springer-Verlag.

Halbach, V. (2014). *Axiomatic Theories of Truth* (revised edition). Cambridge, UK: Cambridge University Press.

Hodges, W. (2008). Tarski's theory of definition. In Patterson, D., editor. *New Essays on Tarski and Philosophy*. Oxford: Oxford University Press., pp. 94–132.

Horsten, L. (1995). The semantical paradoxes, the neutrality of truth, and the neutrality of the minimalist theory of truth. In Cortois, P., editor. *The Many Problems of Realism*. Tilburg: Tilburg University Press, pp. 173–187.

Horsten, L. (2011). *The Tarskian Turn. Deflationism and Axiomatic Truth*. Cambridge, MA: MIT Press.

Horwich, P. (1998). *Truth* (second edition). Oxford: Clarendon Press.

Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford: Oxford University Press.

Ketland, J. (1999). Deflationism and Tarski's paradise. *Mind*, **108**, 69–94.

Kreisel, G. (1967). Informal rigour and completeness proofs. In Lakatos, I., editor. *Problems in the Philosophy of Mathematics*. Amsterdam: North Holland, pp. 138–186.

McGee, V. (1991). *Truth, Vagueness and Paradox*. Indianapolis: Hackett Publishing Company.

Niebergall, K.-G. (2000). On the logic of reducibility: Axioms and examples. *Erkenntnis*, **53**, 27–61.

Pudlák, P. (1998). The lengths of proofs. In Buss, S. R., editor. *Handbook of Proof Theory*, Chapter VIII. Amsterdam: Elsevier Science Publisher, pp. 547–637.

Quine, W. V. O. (1970). *Philosophy of Logic*. Cambridge, MA: Harvard University Press.

Schindler, R. (1994). ACA_0, Π_1-CA_0, and the semantics of arithmetic, and BG, BG + Σ_1-Ind, and the semantics of set theory. Available from www.math.uni-muenster.de/logik/Personen/rds/.

Shapiro, S. (1997). *Philosophy of Mathematics. Structure and Ontology*. Oxford: Oxford University Press.

Shapiro, S. (1998). Proof and truth: Through thick and thin. *The Journal of Philosophy*, **95**, 493–521.

Simpson, S. G. (1999). *Subsystems of Second Order Arithmetic*. Heidelberg: Springer-Verlag.

Suppe, F. (1977). *The Structure of Scientific Theories* (second edition). Illinois: University of Illinois Press

Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica Commentarii Societatis philosophicae Polonorum*, **1**, 261–405.

Väänänen, J. (2001). Second-order logic and the foundations of mathematics. *The Bulletin of Symbolic Logic*, **7**(4), 504–520.

Visser, A. (2006). Categories of theories and interpretations. In Enayat, A., Kalantari, I., and Moniri, M., editors, *Logic in Tehran: Proceedings of the Workshop and Conference on Logic, Algebra, and Arithmetic, (October, 2003)*, Number 26 in Lecture Notes in Logic, pp. 284–341. Boca Raton: A K Peters/CRC Press.

MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY
FAKULTÄT FÜR PHILOSOPHIE
WISSENSCHAFTSTHEORIE

UND RELIGIONSWISSENSCHAFTEN
LMU MÜNCHEN, GESCHWISTER SCHOLL PLATZ 1
D-80539 MÜNCHEN, GERMANY
E-mail: M.Fischer@lrz.uni-muenchen.de

DEPARTMENT OF PHILOSOPHY
43 WOODLAND ROAD
BS83PE, BRISTOL, UK
E-mail: leon.horsten@bristol.ac.uk