THE SUBSTITUTIONAL ANALYSIS OF LOGICAL CONSEQUENCE

Volker Halbach
University of Oxford

19th May 2018

> Consequentia 'formalis' vocatur quae in omnibus terminis valet retenta forma consimili. Vel si vis expresse loqui de vi sermonis, consequentia formalis est cui omnis propositio similis in forma quae formaretur esset bona consequentia [...]
>
> John Buridan, Tractatus de Consequentiis (Hubien 1976, I.3, p.22f)

A substitutional account of logical validity for formal first-order languages is developed and defended against competing accounts such as the model-theoretic definition of validity. Roughly, a substitution instance of a sentence is defined as the result of uniformly substituting nonlogical expressions in the sentence with expressions of the same grammatical category and possibly relativizing quantifiers. In particular, predicate symbols can be replaced with formulae possibly containing additional free variables. A sentence is defined to be logically true iff all its substitution instances are satisfied by all variable assignments. Logical consequence is defined analogously. Satisfaction is taken to be a primitive notion and axiomatized.

For every set-theoretic model in the sense of model theory there exists a corresponding substitutional interpretation in a sense to be specified. Conversely, however, there are substitutional interpretations – in particular the 'intended' interpretation – that lack a model-theoretic counterpart. The substitutional definition of logical validity overcomes the weaknesses of more restrictive accounts of substitutional validity; unlike model-theoretic logical consequence, the substitutional notion is trivially and provably truth preserving. In Kreisel's squeezing argument the formal notion of substitutional validity naturally slots into the place of intuitive validity.

At the origin of logic is the observation that arguments sharing certain forms never have true premisses and a false conclusion. Similarly, all sentences of certain forms are always true. Traditionally, in order to show that an argument is not valid, logicians replaced nonlogical or categorematic terms in an argument with terms of the same respective grammatical category so that the premisses become true and the conclusion false. For instance, the following argument is not valid:

> Some horses are animals. Therefore all horses are animals.

Its validity is refuted by the following substitution instance with a true premiss and a false conclusion:

> Some horses are white. Therefore all horses are white.

From this practice medieval logicians extracted a definition of formal validity: An argument is defined to be *formally* valid if and only if it has no such substitutional counterexamples. Arguably, Buridan was the first to separate this notion of validity explicitly from other notions of validity and to define formal validity explicitly as truth preservation under reinterpretations of nonlogical or categorematic terms.[1]

It is still the case that introductory logic courses often start with a characterization of logical validity as truth preservation under the substitution of nonlogical terms. Together with the informal characterization two properties of logical validity are stressed as essential to logic validity: First, logic is universal. That is, premisses and conclusions can be any declarative sentences of our language and there are no restrictions to a sublanguage.[2] Secondly, logical consequence preserves truth. That is, in a logically valid argument either a premiss is false or the conclusion is true. This property follows directly from the substitutional characterization, because the argument itself, as it stands, would be

[1] Even after the notion of formal validity had been introduced, most logicians followed Aristotle in defining an argument as valid or 'good' if and only if the conclusion follows from the premisses by necessity. Arguments we would call analytically valid were classified as 'good' arguments by logicians, including Buridan. At least from Buridan onwards logicians became clear about the distinction between formal validity and other kinds of validity, although others, including Ockham, used the term *formal* before Buridan, but with a different meaning. See (Read 2012) and (Aho and Yrjönsuuri 2009, §6.3). For the development leading up to Buridan's definition see (Dutilh Novaes 2012). More recently, philosophers have usually relegated the analysis of analytic, metaphysical and other kinds of 'material' validity to philosophy of language and other branches of philosophy. Logical validity has come to be understood as formal validity. See (Asmus and Restall 2012) for a brief historical synopsis. Historically erudite logicians, among them (Read 1994), have argued against the identification of formal and logical validity and the narrowing of the scope of logic to formal validity. Those who prefer this terminology can read 'formal' for my term 'logical' in the rest of the paper.

[2] Of course there may be worries about sentences containing context sensitive vocabulary or anti-realist theories according to which certain declarative sentences are neither true nor false.

a substitutional counterexample, if the premisses were true and the conclusion false. Truth preservation follows from the existence of an intended interpretation; and on the substitutional account, the intended interpretation is the trivial homophonic interpretation that does not modify the argument.

While substitutional elements can still be found in conceptions of logical validity in the Hilbert school, for instance, in (Behmann 1922) and by Hilbert himself, substitutional definitions of validity have been considered obsolete from the middle of the 20th century on by most logicians.[3]

Tarski's work on truth had opened the way to new mathematically precise, nonsubstitutional semantic analyses of logical consequence. These definitions of logical validity follow the same pattern as the substitutional definition in defining validity as truth preservation under all interpretations:[4]

> *Semantic Analysis of Logical Consequence* · An argument is logically valid iff the conclusion is true under all interpretations under which its premisses are true.[5]

In his first attempt *On the Concept of Logical Consequence* (1936*b*) to define logical validity, Tarski applied the theory of satisfaction and truth developed in *The Concept of Truth*

[3] Tarski explicitly rejected such a substitutional analysis of logical consequence as inadequate, because the object language might not provide sufficiently many substitution instances.To take a dramatic example, consider the language of dense linear order without end points. This is a first-order language with one binary predicate symbol R. Now an absolute notion of truth for this language is defined in a language that extends this language by adding a syntax theory and the machinery for carrying out the definition of truth. A sentence ϕ would be defined to follow from a finite set Γ of sentences substitutionally if and only if whenever a way of uniformly substituting Rxy with a formula $\sigma(x, y)$ in Γ and ϕ either makes ϕ true or a sentence in Γ false. By merely using the symbol for the dense linear ordering one cannot obtain countermodels to some invalid arguments. In fact, if the consequence relation were defined this way, the consequence relation would become decidable. Therefore it cannot coincide with the usual consequence relation of predicate logic by Church's theorem. Using the method of quantifier elimination it can be shown that each substitution instance of the argument from Γ to ϕ it is decidable whether all premisses are true and the conclusion false. Again using quantifier elimination, it can be shown that only finitely many substitution instances have to be considered. Hence the substitutional consequence relation defined in this way would be decidable. Thus it cannot coincide with the usual definition of logical consequence for a first-order language with a binary predicate symbol as its only nonlogical symbol.

[4] There are theories of logical consequence that completely deviate from this schema. In the modern era these are mainly the proof-theoretic accounts; but they come from a different lineage, not to be discussed here.

[5] Beall and Restall (2006) called this *Generalized Tarski Thesis*, although they prefer the more general and less specific sounding term *case* over *interpretation*. They credited Jeffrey (1992) with the formulation. Since Tarski has so many achievements to his credit already, I prefer the label above, especially because Tarski himself acknowledged that Bolzano had given a similar definition a century before his work. Tarski added to his (1936*b*) in the English translation in (1956, fn. †). See (Rusnock and Burke 2010) for a comparison of modern accounts of logical consequence and Bolzano's with respect to Etchemendy's (1990) criticism of Tarski's theory.

(1936*a*) to the analysis of logical consequence.[6] The modern model-theoretic definition appeared only in in (Tarski and Vaught 1956) or around that time.

The notion of truth (and satisfaction) needed for one of Tarski's semantic analyses of validity is defined in purely mathematical terms; interpretations are also explained in purely mathematical terms. In this sense both theories are reductionist: All semantic notions are reduced to mathematical notions.

The aim to show that truth is reducible to mathematical terms was one of Tarski's main aims in *The Concept of Truth*. Of course, his theorem on the undefinability of truth imposes restrictions on possible reductions. In his paper on truth, Tarski defined satisfaction and truth in an 'essentially richer' metalanguage. This approach also underlies his analysis of validity in *On the Concept of Logical Consequence* (1936*b*). There, logical consequence is defined for a given object language in an essentially richer metalanguage – or metatheory in modern parlance. As a consequence, logic is no longer universal: Logical consequence is only defined for a fragment of the metalanguage at best (if the object language is contained in the metalanguage). This is the price Tarski had to pay for his reductionist approach to semantics.[7]

The model-theoretic definition of validity has the advantage that universality is retained: The notion of logical consequence is defined for all first-order language in set theory – including the language of set theory itself. There is no need to ascend to a richer metalanguage or stronger theory. However, something else is sacrificed for reductionism: If validity is defined in set theory for the language of set theory, there is no longer an intended interpretation. The interpretations we quantify over in the definition of logical consequence are certain sets with a set-sized domain.[8] No interpretation that interprets the quantifiers as ranging over a set is the intended interpretation. Therefore truth preservation, which is a trivial consequence of the substitutional characterization, is lost.

BACK TO BURIDAN

On both of Tarski's reductionist accounts of logical validity, either the universality of logic or truth preservation of logical consequence is lost. Consequently both definitions have

[6] In (1936*b*) Tarski defined logical truth by first replacing (predicate and individual) constants with variables of the appropriate kind in a sentence of the fixed object language and then defining a sentence as logically valid if and only if is satisfied by all variable assignments for these variables. Tarski called these variable assignments also 'models'. Tarski probably used a fixed domain for all models, although this is at least contentious. See (Bays 2001) and the references therein.

[7] Recently, (Williamson 1999*a*, 2013) defended an analysis of validity in the style of Tarski's (1936*b*) account, although many philosophers remain sceptical, especially after Etchemendy's (Etchemendy 1990) attack on Tarski's *On the Concept of Logical Consequence*.

[8] The reason for the restriction to sets as ranges of quantifiers is again reductionism: The inductive definition of satisfaction can only be shown ot have a fixed point if the domain is a set.

come under attack as adequate analyses of logical validity.[9] Since both, universality and truth preservation, are so central to our concept of logical validity, it may be promising to revisit the old substitutional characterization.

I start from the general form of Semantic Analysis of Logical Consequence given above. The reductivist approach forced Tarski either to relativize the notion of truth – and, consequently, logical consequence – to an object language or to set-theoretic models. To avoid these restrictions, I abandon reductivism and employ a primitive axiomatized notion of satisfaction.[10] Using the primitive notion of satisfaction, a sentence will be defined as logically valid if and only if all its substitution instances are satisfied. Logical consequence is defined analogously. In what follows I will often just talk about logical validity of single sentences, because this makes the presentation slightly easier; but everything is intended to apply to logical consequence *mutatis mutandis*.

There are various reasons why I expect a substitutional definition of logical validity defined from a primitive notion of satisfaction to be preferable to the usual model-theoretic definition. First, the substitutional definition of validity is closer to traditional informal characterizations of validity as they are given in introductory logic courses and used in philosophy outside logic. It is also closer to how logicians over the centuries specified counterexamples and established validity before the model-theoretic account finally became prevalent. In this sense it is more direct and deviates less from the traditional notion.

Secondly, on a substitutional account it is obvious why logical truth implies truth *simpliciter* and why logical consequence is truth preserving: If a sentence is logically valid, it is true under all substitutional interpretations. The 'intended interpretation' is one of these interpretations, and thus logical validity implies truth. Moreover, in contrast to the model-theoretic account, there is nothing mysterious about the intended interpretation on the substitutional account: It is the homophonic interpretation that takes sentences at their face value; it does not substitute anything and maps any sentence to itself as a substitution instance.[11]

Thirdly, the substitutional definition of logical consequence is not tied to set theory and

---

[9] Etchemendy (1990) has established the analysis of logical consequence again as a widely discussed topic in philosophy of logic, although many of his points can be found in earlier work, for instance, in (Kreisel 1967*a*).

[10] In this respect the present approach differs from earlier mathematically presented substitutional theories of consequence such as Quine's (1970). Reductionism causes there problems similar to those discussed above. For a further discussion of Quine's theory see (Eder 2016).

[11] Under Linnebo's (2012) definition logical validity is also obviously truth preserving. Linnebo's approach resembles the approach defended here in various aspects. However, he does not employ a primitive satisfaction predicate but rather a series of property theories. Instead of substitutional interpretations Linnebo employs interpretations using properties. In many respects this can be taken to be a formal elaboration of Bolzano's theory mentioned above. The substitutional interpretations used in appendix 2 in the proof of substitutional completeness bear a certain resemblance with Linnebo's interpetations, just with satisfaction instead of property instantiation.

its philosophy. On the model-theoretic account, interpretations are specific sets; on the substitutional account they are merely syntactic and (under certain natural assumptions) computable functions replacing expressions and perhaps an assignment of objects to the free variables. The quest for an adequate model-theoretic conceptual analysis of logical validity has led philosophers to doubt that quantification 'over absolutely everything' is possible and to speculate about the indefinite extensibility of the set-theoretic universe. On a substitutional account it is easier to avoid such speculations. At any rate the direct link between the most complex metaphysical speculative theory hitherto, set theory, and the theory of logical consequence is severed. This does not mean that I reject set theory; I just would like to avoid making the theory of logical consequence dependent on the philosophy of set theory

Before discussing the notions of satisfaction and a substitution instance, I fix some assumptions on the language and the mathematical framework in and for which the theory is going to be developed. It is one of the main aims of this paper to demonstrate how the substitutional account can overcome the problems of the model-theoretic account. To facilitate the comparison with model theory, I define substitutional and model-theoretic consequence in the same background theory. Model theory is usually carried out in a set theory. Consequently, I define substitutional consequence in a theory extending first-order set theory, that is, Zermelo–Fraenkel or some variant thereof.

The substitutional account could also be developed in a much weaker account. Only some syntactic notions are required to define substitution instances. However, it would be difficult to compare this substitutional analysis with the model-theoretic definition. Moreover, the goal is to obtain a universal notion of logical consequence that pertains to a comprehensive language, for instance, an idealized version of our overall language of mathematics or science. I do not see any good reason to exclude set theory from it. Investigating the substitutional definition of logical consequence in a weak theory of syntax or arithmetic may be a worthwhile exercise, but is not conducive to the purposes of this paper.

SUBSTITUTION INSTANCES

In this and the next section the two key notions in the substitutional analysis of logical truth and consequence are made formally precise. First, I turn to the notion of a substitution instance and in the next section with truth.

A particular substitution instance of a formula or an entire argument is given by a *substitution function*. A substitution function uniformly replaces all nonlogical terms with expressions of the same grammatical category in any given formula. Hence the definition of a substitution function will rely on the distinction between logical and nonlogical terms. I adopt the usual distinction in first-order predicate logic between logical and nonlogical terms: Connectives, quantifiers, and identity are taken as logical terms, while predicate symbols (and function symbols, including constants, if present) are conceived as nonlogical. Here I do not provide a justification for making the distinction in this particular way. Other choices of logical constants could be accommodated. For instance, it would not be a problem to treat identity as a nonlogical symbol or another predicate symbol as a logical constant.

Even once the logical terms are fixed, it may not be clear what counts as a suitable substitution instance. To motivate the definition for formalized languages, I look at examples in natural language. What would count as a substitutional counterexample to the following argument?

> Some horses are animals. Therefore all horses are animals.

Clearly, a general term such as 'is an animal' can be replaced not only with another general term but also with a complex general term such as 'is a European country adjacent to Austria'. Similarly, it is permissible to replace atomic singular terms with complex ones. Replacing proper names with definite descriptions may cause problems; and this requires some care. I omit the elaboration of the formal details.

There are more singular terms than only proper names and definite descriptions. Should personal or demonstrative pronouns be admitted as substitution instances of proper names? As an example I consider the following invalid argument:

> Some horses are animals. Bucephalus is a horse. Therefore Bucephalus is an animal.

Would the following be a substitution instance and a counterexample, assuming the speaker points at some unnamed spider?

> Some animals are horses. That is an animal. Therefore that is a horse.

'Horse' is replaced with 'animal' and vice versa; the pronoun name 'that' is substituted for the name 'Bucephalus'. I cannot see any reason to reject this counterexample; it is also in line with counterexamples in traditional logic.[12] Of course, the reference of pronouns must be kept fixed between premisses and the conclusion. In the example both occurrences of 'that' must refer to the same object. With that proviso pronouns can be substituted for singular terms. This applies at least to personal and demonstrative pronouns.

Pronouns can also be introduced via complex predicate expressions. For instance, the predicate expression 'is a horse' may be substituted with 'is smaller than this but bigger than that'. Since occurrences of pronouns in substitution instances are going to be allowed, the definition of logical truth in natural language would require a reference to the way the pronouns are interpreted:

[12] Already Aristotle used demonstrative pronouns in syllogisms, for instance, in the Nichomachian Ethics VII 3. Many further examples can be found throughout history.

> A sentence is logically true iff all substitution instances are true for any reference of the pronouns.

The use of demonstrative or personal pronouns makes it possible to formulate counterexamples or interpretations involving a singular term referring to an object for which we lack a name or definite description. In the formal language free variables will play the role of these pronouns. Thus a substitution instance of a sentence may contain free variables.

Now assume some first-order language is fixed. In a nutshell, a substitution function uniformly replaces each predicate symbol with some formula and possibly relativizes quantifiers.[13] More precisely, the notion of a substitution function is defined inductively. To keep the presentation simple, I assume that the language does not contain any individual constants or function symbols. They can be dealt with in well-known ways. The clauses of the inductive definition can be sketched as follows: If I is a substitution function and ϕ an atomic formula, then $I(\phi)$ is some formula containing at least the same free variables as ϕ itself. This holds even if ϕ contains the satisfaction predicate. Only for identity we do make an exception, as it is to be treated as a logical constant: The identity symbol is not replaced by any substitution function. Substitution functions commute with the connectives, that is, $I(\neg\phi)$ is $\neg I(\phi)$ and $I(\phi \wedge \psi)$ is $I(\phi) \wedge I(\psi)$ for any formulae ϕ and ψ, and similarly for further connectives. Substitution functions can treat quantifiers in two different ways. A substitution function is unrelativized if $I(\forall x\, \phi)$ is $\forall x\, I(\phi)$. A substitution is relativized if there is a formula $\delta(x)$ such that $I(\forall x\, \phi)$ is $\forall x\, (\delta(x) \rightarrow I(\phi))$ for all formulae ϕ. The formula $\delta(x)$ serves the purpose of the domain on the model-theoretic account. For the moment being, both, relativized and unrelativized substitution function are taken into account.[14] The clauses of the definition are not quite correct: Variables may be bound inadvertently in $\delta(x)$ or the translation of atomic formulae. To avoid these variable clashes variables have to be renamed in a suitably way. I skip the cumbersome details.

SATISFACTION AS A PRIMITIVE NOTION

For the analysis of logical validity advocated here, semantic reductionism has to be abandoned. Since Davidson's work on truth in the 1960s, philosophers have slowly warmed to the idea of taking truth or satisfaction as a primitive notion.[15] In this paper the axioms

[13] Substitution functions resemble relative interpretations as introduced by Tarski et al. (1953). Substitution functions are defined like relative interpretations with the same theory as source and target theory; only the requirement that provability is preserved is dropped.

[14] It may be objected that relativized substitution function do not treat quantifiers as logical constants. Of course a similar objection can be made against the use of a domain in model-theoretic semantics. In contrast to model-theoretic semantics, quantifiers can easily be treated as strict logical constants by ruling out relativized substitution functions. I admit them here in order to facilitate a comparison with the model-theoretic definition of logical consequence.

[15] Tarski (1936*a*) already considered axiomatic approaches. Davidson advocated an axiomatic approach in various papers (see his 1984) with satisfaction as a primitive notion. Asay (2013) provided a defence

for satisfaction will be added to an extension of Zermelo–Fraenkel set theory possibly with urelements as the 'base theory'. The base theory can be enriched with further defined notions; also further axioms may be added. The base theory can be taken our overall theory. Starting from a comprehensive theory is important for the universality of logic. It is not needed from the technical point.

The binary satisfaction predicate $\mathrm{Sat}(x, y)$ is intended to apply to formulae x and variable assignments y. Variable assignments are functions from the set of variables (or their indices); there is no restriction on the range of variable assignment. The schemata of Zermelo–Fraenkel are expanded to the full language including the satisfaction predicate. The entire theory including formulae involving Sat is governed by classical logic.

The substitutional theory of logical validity requires only weak axioms for satisfaction. In appendix 1 the axioms needed are listed as 'mandatory' axioms. Here in the main body of the paper I merely sketch these mandatory axioms. For atomic formulae of set theory the following axiom is stipulated: A variable assignment a satisfies the formula $x \in y$ if and only if $a(x) \in a(y)$, where $a(v)$ is the value of a given variable v under the variable assignment a. Similar axioms are added for all predicate symbols other than Sat. If individual constants and function symbols are present, suitable axioms have to be specified. Compositional axioms are added stating that the satisfaction predicate commutes with connectives and quantifiers. Hence there is an axiom expressing that a variable assignment satisfies a formula $\phi \wedge \psi$ if and only if it satisfies ϕ and ψ; a variable assignment satisfies a formula $\neg\phi$ if and only if it does not satisfy ϕ; and so on for other connectives. A variable assignment satisfies a formula $\forall x\, \phi$ if and only if all its x-variants satisfy ϕ. As usual, an x-variant of a variable assignment is any variable assignment that differs from it only in the value of x. The formulae and sentences ϕ and ψ may contain the satisfaction predicate.

I use the name Ω for the overall theory with axioms for set membership and satisfaction. The theory Ω resembles the usual 'Tarskian' theory of truth – with the exception that the compositional clauses apply also to formulae containing the satisfaction predicate.

Since schemata (separation and replacement) are also expanded to the language with the satisfaction predicate, Ω is properly stronger than the base theory. Its consistency cannot be proved relative to Zermelo–Fraenkel set theory. However, it is equiconsistent with a typed Tarskian theory and thus there are good reasons to believe in its consistency if the base theory is assumed to be consistent. Moreover, adding analogous axioms to reasonably behaved weaker theories such as Peano arithmetic yields provably consistent extensions of these theories. The truth axioms act as a reflection principle.[16] Therefore it is at least plausible to assume the consistency of Ω.

The axioms for satisfaction describe a notion of truth based on classical logic: Satisfac-

of primitivism about truth. Halbach (2014) gave a survey of axiomatic theories of truth mainly for arithmetical languages; Fujimoto (2012) investigated truth theories over set theory.

[16] See (Halbach 2014, sec. 22).

tion commutes with all connectives and quantifiers.[17] It reflects the axioms for classical logic in which the base theory is formulated. In principle the substitutional theory of logical consequence is also compatible with nonclassical logic. If a nonclassical theory were used as base theory and if a theory of validity in some nonclassical logic were the goal, the axioms for satisfaction would have to be adjusted accordingly.

Substitutional definitions of logical truth & consequence

A formula ϕ is said to be satisfied under a substitution function I and a variable assignment iff the substitution instance $I(\phi)$ of that formula is satisfied under that variable assignment. Using the notions of satisfaction and substitution function, logical truth and consequence are defined in the theory Ω:

> *Substitutional Definition of Logical Consequence* · A sentence ϕ is a logical consequence of a premiss set Γ iff ϕ is satisfied under all substitution functions and all variable assignments under which all formulae in $\Gamma \cup \{\exists x\, x = x\}$ are satisfied.

The variable assignments are needed in the definition, because a substitution instance of a formula may contain free variables, even if the premisses and the conclusion do not.

Logical truth is defined from the substitutional notion of consequence in the usual way: A sentence is logically true if and only if it is a logical consequence of the empty premiss set. Therefore, a sentence is logically true iff it is satisfied under all substitution functions and variable assignments that satisfy $\exists x\, x = x$.

If the extra premiss $\exists x\, x = x$ were dropped in the definition of logical consequence, a notion of logical consequence would be obtained that is ontologically neutral and the claim that there is at least one object would not be logically true. The reason is that a substitution function may relativize all quantifiers to a formula $\delta(x)$ that is not satisfied by any object. For instance, nothing in the definition of a relativized substitution function rules out $x \neq x$ as relativizing formula. In order to force the extensional equivalence with the usual model-theoretic definition, I add the extra premiss in the substitutional definition. If the substitution function I is relativizing quantifiers to $\delta(x)$, then the substitution instance of $\exists x\, x = x$ is $\exists x\, (\delta(x) \wedge x = x)$, which is not satisfied if $\delta(x)$ is not satisfied by anything.[18]

[17] Some popular type-free theories of truth such as the Kripke–Feferman theory (Reinhardt 1986, Feferman 1991) are not based on classical logic: although they are formulated in classical logic, they describe a nonclassical concept of truth. Typically, truth doe snot commute with negation in these theories.

[18] Here I consider only languages without individual constants or function terms. If they are admitted, the definition of a substitution function can be modified in various ways so that a consequence relation in some version of free logic is obtained. The substitutional account of logical consequence yields a smooth treatment of free logic. Here I do not go into details and concentrate on the variety that coincides with classical consequence.

This does not mean that the definition with the additional premiss $\exists x\, x = x$ is superior. In fact, there are good reasons to omit it. Here it is only included because it facilitates the comparison with the model-theoretic account. It merely accommodates a quirk of the classical model-theoretic definition, where domains of models are not allowed to be empty. I suspect that the reason for this restriction in model theory is mainly technical: If the domain of a model is empty, then, trivially, there is no variable assignment over that domain. Hence, if truth is defined as satisfaction by all variable assignments, every sentence is true over the empty domain, even propositional contradictions.[19] On the model-theoretic account it is awkward to stay ontologically neutral; on the substitutional account it is natural and straightforward to omit the extra premiss $\exists x\, x = x$.

There are no restrictions on the sentences in the definition of logical consequence. In particular, the sentence ϕ and the sentences in Γ may contain the satisfaction predicate. Like the model-theoretic concept of consequence and unlike Tarski's (1936*b*) analysis, the substitutional definition does not require the ascent to a stronger metatheory. Logical consequence is defined in Ω for the language of Ω. In this sense the substitutional conception of logical consequence is universal in the same way the modern model-theoretic conception is universal. It is also absolute – like Tarski's (1936*b*) and unlike the model-theoretic analysis – because the notion of satisfaction is absolute and not relativized to a model or interpretation.

With this definition it is trivial that logical validity implies truth. If a sentence is logically valid, it is satisfied under all substitution functions and variable assignments. The identity function, that is, the function that maps every formula to itself is a substitution function. Hence, if a sentence is valid on the substitutional definition, it is satisfied under all variable assignments, that is, it is true *simpliciter*. Similarly, logical consequence preserves truth. The argument for the claim that logical consequence preserves truth follows the traditional 'naive' informal argument that is also given in introductory logic classes. This is in contrast with the model-theoretic account.

The argument that logical truth implies truth can be explained more formally as follows. Assume that $\mathrm{Val}(x)$ is the formula in Ω defining logical truth in the substitutional sense. Then the claim that logical validity implies truth becomes $\forall x\,(\mathrm{Val}(x) \rightarrow \forall a\, \mathrm{Sat}(x,a))$, where $\forall a$ expresses quantification over all variable assignments (the restriction of the quantifier $\forall x$ to sentences is not necessary). This principle is not even expressible on the model-theoretic account, because there is no absolute set-theoretically definable satisfaction predicate Sat.

It might be objected that $\forall x\,(\mathrm{Val}(x) \rightarrow \forall a\, \mathrm{Sat}(x,a))$ does not express what was intended by saying that logical validity implies truth. What was really meant, so it might

[19] As Schneider (1958) observed, if a sentence is defined to be true iff it is satisfied by all variable assignments, then all sentences will be true in the model with the empty domain, because there are no variable assignments over the empty domain. Of course, the problems can be solved in the model-theoretic framework. Williamson (1999*b*) discussed some workarounds; but for most purposes it is much more convenient just to exclude the empty domain.

be claimed, is the schema $\mathrm{Val}(\ulcorner\phi\urcorner) \to \phi$ for all sentences ϕ, where $\ulcorner\phi\urcorner$ is some canonical name for the sentence ϕ. In the absence of an absolute truth or satisfaction predicate as on the model theoretic account, this may be the best approximation; but it is at best only an approximation. It is not only a problem that the principle that validity implies truth has to be expressed by an infinite set of sentence; there is also a deeper hidden problem, if this schema is employed. The reasoning for showing that validity implies truth should proceed in the 'naive' way as follows: If a sentence is true under all interpretations (whatever they might be), then it is true under the intended interpretation and therefore true *simpliciter*. Whether the truth of the sentence implies the sentence itself, is then another matter: In the case of sentences without the truth or satisfaction predicate and for many other sentences, this step is sound; but in the case of liar sentences and the like this is far from obvious. In fact, it is not licensed by many popular truth theories and my preferred theory sketched in appendix 1.[20]

THE SQUEEZING ARGUMENT

The model-theoretic and the substitutional definition of logical truth have the same extension. I sketch a proof for this claim. In fact, I show something much stronger: I show that the notion of substitutional validity naturally slots into the place of Kreisel's (1967*a*) 'intuitive validity' in his squeezing argument.[21]

Kreisel (1967*a*) distinguished intuitive validity from truth in all set-theoretic structures, that is, model-theoretic validity. 'Intuitive validity' is not to be understood as some kind of pre-theoretic notion of logical truth or the like. In fact, I doubt that there is a pre-theoretic notion of logical validity. Logical validity is a highly theoretical notion, even though the theorizing began in antiquity. Kreisel thought that intuitive validity is a rigorous but informal notion. By 'informal', it seems, he meant 'not mathematically defined' (see Smith 2011).

According to the squeezing argument, provability of a sentence in a suitable deductive system such as Natural Deduction ('$\vdash_{\mathrm{PC}} \phi$' in the diagram below) implies its intuitive validity because of the 'intuitive' soundness of the chosen calculus. The intuitive validity of

[20] It can still be asked whether the schema $\mathrm{Val}(\ulcorner\phi\urcorner) \to \phi$ is provable in Ω. In fact it is, but perhaps, the critical reader might object, for the wrong reasons. The proof is somewhat convoluted. First it can be shown that $\mathrm{Val}(\ulcorner\phi\urcorner)$ implies truth in all set-theoretic models and hence provability in predicate logic by the formalized completeness theorem. The theory of satisfaction is essentially reflexive and thus proves the local reflection principle for any finite subtheory, and thus for logic. A proof via the satisfaction predicate is not feasible, even if the theory of satisfaction is (consistently) strengthened. The schema $\forall a\, \mathrm{Sat}(\ulcorner\phi\urcorner, a) \to \phi$ is inconsistent, as Sat commutes with negation. Only the *rule* that licenses the step from a proof of $\forall a\, \mathrm{Sat}(\ulcorner\phi\urcorner, a)$ to ϕ can consistently be added. (Kreisel 1965, p. 117) gave a similar argument, pointing out that it does not apply to finitely axiomatized theories.

[21] Kreisel presented the squeezing argument in a number of places with variations. (Kreisel 1967*a*, p. 152–157) is the standard reference. I am indebted to Göran Sundholm for making me aware of (1965, p. 116f.) and (1967*b*, p. 253ff.). Beau Mount brought the relevant paragraph in (1980, p. 177) to my attention.

a sentence in turn implies its model-theoretic validity (symbolized as '$\vDash$' below), because any model-theoretic counterexample is also an 'intuitive' counterexample and thus refutes the intuitive logical truth of a sentence. These two implications – both straight arrows in the diagram below – cannot be proved formally, because the notion of intuitive validity is not formal. However, there is a formal theorem, Gödel's completeness theorem for first-order logic, that shows that the model-theoretic validity of a sentence implies its provability in the chosen logical calculus. The implications are shown in the following diagram:

$$\begin{tikzcd}[column sep=huge]
\vdash_{\mathrm{PC}} \phi \arrow[r, "\text{intuitive}", "\text{soundness}"'] & \begin{array}{l}\phi \text{ is}\\ \text{intuitively}\\ \text{valid}\end{array} \arrow[r, "\text{every countermodel}", "\text{is a counterexample}"'] & \vDash \phi \arrow[ll, bend left, "\text{completeness theorem}"]
\end{tikzcd}$$

If the two informal implications visualized by the two straight arrows in the diagram hold, all three notions have the same extension. For the mathematical logician, fixing the extension of logical validity is usually good enough. But the squeezing argument does not establish that either of the two formal definitions is an adequate conceptual analysis of logical validity.

If intuitive validity is replaced with substitutional validity, all implications become formally provable and a *formally* rigorous version of the squeezing argument is obtained:

$$\begin{tikzcd}[column sep=huge]
\vdash_{\mathrm{PC}} \phi \arrow[r, "\text{formal}", "\text{soundness}"'] & \begin{array}{l}\phi \text{ is}\\ \text{substitutionally}\\ \text{valid}\end{array} \arrow[r, "\text{substitutional}", "\text{completeness}"'] & \vDash \phi \arrow[ll, bend left, "\text{completeness theorem}"]
\end{tikzcd}$$

The arrow labelled 'formal soundness' is provable using the axioms for satisfaction. The proof now resembles the kind of justification for the rules and axioms of the chosen calculus given in introductory logic courses. For instance, we may say that the rule of 'or'-introduction in Natural Deduction is sound, because, whenever a sentence is true, the disjunction of that sentence with a sentence is true as well. This holds for all substitution instances of that rule.

To establish the implication marked 'substitutional completeness' I show a much stronger result resembling the informal argument given for the corresponding implication in the original squeezing argument. For any given set-theoretic model an equivalent substitutional interpretation, that is a pair consisting in a substitution function and a variable assignment, will be specified in a straightforward way. Thus, any model-theoretic counterexample is also a substitutional counterexample (modulo the trivial transformation).

The point of this observation is not only that it establishes the formally rigorous squeezing argument, but also that the main worry about the substitutional approach does not apply to the version in this paper: By passing from set-theoretic validity to substitutional validity no interpretations are lost. In fact, there are 'more' substitutional interpretations than there are set-theoretic models: For any given set-theoretic model there is a corresponding substitutional interpretation, but there are substitutional interpretations that do not have a model-theoretic counterpart, the intended interpretation being an example.

The completeness result can be stated in the following form:

> *Substitutional Completeness* · In Ω the following statement is provable: For any model M there is a substitution function I and a variable assignment a such that the following equivalence holds for all sentences ϕ :
>
> $$M \vDash \phi \text{ iff } \mathrm{Sat}(I(\phi), a)$$

The proof is sketched in appendix 2. The claim can be strengthened in various ways. As will be obvious from the proof, the same substitution function I can be used for all models; only the variable assignment has to be varied. The claim can also be generalized to formulae ϕ with free variables.[22] Moreover, not all axioms for satisfaction are needed in the proof.[23]

With the Substitutional Completeness theorem all implications in the formally rigorous squeezing argument have been established. The theorem shows that all three notions, provability, model-theoretic validity, and substitutional validity coincide in their extensions. Hence the Substitutional Completeness theorem can be taken as an extensional reduction of the semantic concept of logical consequence to the purely mathematically defined notion of model-theoretic validity and the purely syntactic concept of provability.

ROBUSTNESS

I will now address some objections that have been made against versions of the substitutional account. Many of them are founded in the suspicion that substitutional definitions make the notion of logical validity highly language-dependent. As I mentioned above, Tarski (1936*b*) rejected a substitutional definition because it can be extensionally inadequate if the object language is expressively weak.[24] However, this objection is based

[22] The generalization can be formulated in the following way: ϕ holds in M under the variable assignment a over M iff $\mathrm{Sat}(I(\phi), a)$. Variables in ϕ' may have to be renamed and a be adjusted accordingly to avoid variable clashes. This generalization shows that the substitutional interpretation $\langle I, a\rangle$ is in a way really the 'same model' as M: They are elementarily equivalent even with all parameters from M added.

[23] First, the uniform typed disquotation schema is needed; that is, the satisfaction predicate needs to be applied only to formulae not containing the satisfaction predicate. Secondly, the satisfaction predicate needs to commute with the connectives schematically. The universally quantified versions of the mandatory axioms listed in appendix 1 are not needed.

[24] Etchemendy discussed a closely related problem, the persistence problem, in various places in his (1990).

on a strict distinction between object and metatheory, which in turn has its origin in Tarski's reductive approach to semantics. In fact, my substitutional definition can be applied to highly restricted languages, for instance, the language of dense linear order without end points, which has been mentioned above in footnote. 3. Of course, Tarski was right that the substitutional approach yields an incorrect definition if only substitution instances from such a restricted language are admitted in the definition of logical truth and consequence. But if this restriction is lifted and arbitrary substitution instances are admitted, the problems identified by Tarski disappear. This requires a type-free notion of truth that is not restricted to some weak object language. Such a notion of truth or satisfaction is unavailable to the semantic reductionist, including Tarski.

A related worry is that the definition of logical validity is crucially dependent on the chosen base language, that is, the language from which substitution instances are taken. The definition advanced here, however, is highly stable under variations of the language. The formally rigorous squeezing argument and the Substitutional Completeness theorem go through with only one binary predicate symbol for set membership.

One might now suspect that the substitutional analysis is too sensitive to the base *theory* rather than the language. After all, I advertized my substitutional definition as a precise version of the traditional account dating back at least to Buridan, but then used set theory as a base theory, although that was hardly the background on which traditional logicians worked. As I explained above, I chose set theory mainly because it facilitates the comparison with the model-theoretic analysis. The substitutional theory could also be developed in a much weaker base theory. What will be needed is a theory of syntax. Without syntax theory it is not possible to define the notion of a substitution function and, in fact, to state the substitutional definition of validity. It also supplies countably many objects that can be assigned to variables. Hence. as soon as a decent theory of syntax is assumed, the Gödel completeness theorem and the formally rigorous squeezing argument will become provable, although the Substitutional Completeness theorem with respect to model theory will not be provable in the form given above.[25] At any rate, the substitutional theory is highly independent of the chosen base theory and can be developed using only a basic theory of syntax. The model-theoretic approach, in contrast, is much more reliant on set theory, even though analogies of model-theoretic results can be formulated and proved in much weaker theories.

For the substitutional theory, I had to invoke a primitive notion of satisfaction given by axioms. If the substitutional account were highly sensitive to the specific axioms of a type-free theory of truth or satisfaction, any worries about the solution of the paradoxes in

[25]If the base theory is a syntax theory or some arithmetical theory, the axioms for satisfaction will have to be modified, because variable assignments cannot be treated any longer as infinite sequences. In the traditional account more than mere syntax theory was available. I suspect that analogues of the Substitutional Completeness theorem may be available within a framework of traditional metaphysics. A notion of property instantiation or the theory of satisfaction will always interpret a theory of higher-order objects.

the chosen axiomatic theory of satisfaction would also affect the theory of logical validity. As pointed out before, the axioms needed for satisfaction are very weak. When it comes to applications of the satisfaction predicate to formulae containing the satisfaction predicate, I merely require that satisfaction commutes with connectives and quantifiers. Beyond these axioms, only a disquotation principle for satisfaction for formulae without the satisfaction predicate is needed. No further assumptions on the treatment of formulae with the satisfaction predicate have to be made. Even if a typed theory of satisfaction were chosen, much of the above would not be affected. The notion of validity could still be universal and apply to sentences containing the satisfaction predicate; but a sentence would be defined as logically valid iff all its substitution instances in the language without the satisfaction predicate were satisfied by all assignments. The formally rigorous squeezing argument could still be proved. However, the intended interpretation conceived as the substitution function that maps any formula to itself, would no longer be a substitution function in the new sense, because substitution functions would not be allowed to have formulae containing the Sat-predicate as outputs. This restriction strikes me as unmotivated and unnecessary; moreover one would require a squeezing argument showing that the exclusion of formulae with Sat does not affect the extension of the notion of substitutional logical validity.

Why should one employ a primitive notion of satisfaction at all? Why should one not directly use a primitive notion of logical validity, if set-theoretic reductionism is abandoned and semantic notions including satisfaction and logical consequence do not have to be reduced to purely mathematical notions?[26] There are good reasons against taking logical consequence rather than satisfaction (or both notions) as primitive. First, we cannot easily appeal to a intuitive pretheoretic notion of logical consequence. The notion of logical consequence is theoretical, as remarked above, and usually introduced by a characterization in the style of the Semantic Analysis of Logical Consequence above. Truth and satisfaction are usually taken to be conceptually prior to logical consequence. Secondly, as logical truth and consequence can be defined from a theory of satisfaction using only syntactic notions, we should minimize the number of primitive notions by defining validity from satisfaction. Conversely, satisfaction cannot be defined in terms of logical validity and set theory. If logical validity is treated as a primitive notion, truth or satisfaction still would have to be added as a primitive notion, in order to show truth preservation and other desired properties of logical validity. That satisfaction and truth are not definable for first-order languages in terms of logical validity is due to recursion-theoretic grounds: Logical validity will be extensionally equivalent to first-order provability and thus be recursively enumerable, while truth is not even elementarily definable. Therefore

[26] Recently Field (2015) argued that validity should be taken as a primitive. There are much earlier examples of the treatment of validity as a primitive axiomatized notion. In particular, (Kreisel 1965, 1.83) pursued such an approach, but then uses his squeezing argument to show that the axiomatized notion of validity is reducible to set theory. He also considers a variant in 1.84 with a validity predicate that is not reducible to set theory if the underlying theory is finitely axiomatizable.

primitivism about satisfaction is preferable to primitivism about logical validity (or both, validity and satisfaction).

INEXPRESSIBLE INTERPRETATIONS

On the model-theoretic account of logical validity, the quantification over interpretations in the definition omits the intended interpretation; in the substitutional definition, the intended interpretation it is included. However, it may be suspected that, still, the substitutional definition suffers in the end from similar problems as the model-theoretic definition, because there are 'intuitive interpretations' that do not correspond to any substitutional interpretation. 'Intuitive interpretations' would be different from substitutional and model-theoretic interpretations. If there are such elusive intuitive interpretations, then the substitutional analysis of logical validity shares the fate of the model-theoretic definition: It is at best extensionally correct, but not an intensionally adequate analysis.

How could one argue for the existence of such intuitive interpretations that cannot be viewed as substitutional interpretations and do not have a substitutional counterparts? They could be understood as structures similar to model-theoretic structures, but not necessarily limited by any cardinality restrictions; they could be proper classes or pluralities of some sort.[27] Of course, all set-sized models can be rewritten as substitutional interpretations by the Substitutional Completeness theorem. Hence these intuitive interpretations would have to be properly class-sized. One could try to prove the existence by showing the existence of such structures in a class theory. However, von Neumann–Gödel–Bernays set theory will not suffice. This theory only proves the existence of all predicatively definable classes; but all predicatively definable structures can be substitutionally expressed.[28] An impredicative class theory such as Morse–Kelley set theory will suffice.

However, if Morse–Kelley is accepted and used to demonstrate a problem for the substitutional theory, then the base theory should be revised. The base theory should now contain our overall mathematical theory. Thus, if Morse–Kelley is accepted, Zermelo–Fraenkel should be replaced with Morse–Kelley as base theory. It is no problem to adapt the axioms for satisfaction to a theory like Morse–Kelley. If one claims that there is a class, a set, a plurality M of some kind that is an interpretation that cannot be expressed substitutionally, then that plurality can be used again to construct a substitutional interpretation plus a variable assignment that corresponds to this M in the extended theory Ω over the new base theory. The Substitutional Completeness theorem can be proved for a class theory as base theory, and thus there is a substitutional interpretation for any

[27] I thank an anonymous referee and some colleagues for pressing that point.

[28] To see this, one can relatively interpret von Neumann–Gödel–Bernays set theory in Ω by making use of the satisfaction axioms. This interpetation is very similar to the interpretation of the theory ACA of arithmetical comprehension in the Tarskian theory CT of truth over Peano arithmetic. For details see (Halbach 2014, 8.6).

structure in Morse–Kelley. Something similar may also be possible for plural quantifiers of similar devices.

In an attempt to salvage the criticism by going modal, it could be claimed that there *could* be further sets and structures that do not actually exist. I find it very hard to make sense of the modality if pure sets are concerned, even though some philosophers working on the indefinite extensibility of the set-theoretic universe are more optimistic. But, again, if one were serious about possible but not actual intuitive interpretations, one would have to inject the modal principles into the base theory. It is then not trivial to formulate the axioms for satisfaction for such a contingentist modal language.[29] I would expect that an argument analogous to proof of the Substitutional Completeness theorem still can be proved.

CONCLUSION

For first-order logic, the notions of provability, model-theoretic, and substitutional validity coincide. For the purposes of mathematical logic one can use whatever notion comes in handy. As a conceptual analysis of logical consequence, the substitutional notion has some advantages: First, it is a notion of consequence that is universal, just like the model-theoretic notion, in the sense that it applies to the entire language and is not confined to some sublanguage. Secondly, unlike the model-theoretic definition, it makes truth preservation a trivial property of logical consequence. This shows that the pre-model-theoretic 'naive' reasoning about logical consequence can be made formally precise and that there is nothing wrong with accepting that logical consequence is truth preserving and universal at the same time. The price that has to be paid is the rejection of semantic reductionism: The substitutional account requires a primitive notion of satisfaction that cannot be eliminated by purely mathematical concepts such as set membership.

APPENDIX 1: THE AXIOMS FOR SATISFACTION

The following axioms on Sat are used in the proof of Substitutional Completeness. They are also needed to prove the soundness of the chosen calculus, that is, for the proof that ϕ follows substitutionally from Γ if $\Gamma \vdash \phi$. T1 is the schema of 'uniform' disquotation for Sat-free instances. The axioms T2–T4 state that Sat commutes with connectives and quantifiers for all sentences, including those with Sat. The axiom schemata of set theory are extended to the language with Sat.

I introduce some abbreviations. The quantifiers $\forall\phi$ and $\forall\psi$ range over (the codes of) all formulae of the full language with Sat. This can be expressed using a suitable formula defining the set of formulae in the language of set theory. The quantifiers $\forall v$ and $\forall w$ range over all variables. The quantifiers $\forall a$ and $\forall b$ range over variable assignments, that

[29] This point is related to Etchemendy's contingency problem see (Etchemendy 1990) and (McGee 1992).

is, arbitrary functions from the set of variables. $a(v)$ is the value of v under a. Of course all these operations and syntactic operations such as the function that yields applied to a formula its negation need to be expressed in the language of set theory. The reader is referred to (Halbach 2014) for details. For truth theories over Zermelo–Fraenkel in general set theory see (Fujimoto 2012).

Mandatory Axioms

The following axioms for the satisfaction predicate are required:

- T1 $\forall a\, \forall v\, \forall w\, \big(\mathrm{Sat}(\ulcorner Rvw \urcorner, a) \leftrightarrow Ra(v)a(w)\big)$
  and similarly for predicate symbols other than Sat
- T2 $\forall a\, \forall \phi\, (\mathrm{Sat}(\ulcorner \neg\phi \urcorner, a) \leftrightarrow \neg\mathrm{Sat}(\ulcorner \phi \urcorner, a))$
- T3 $\forall a\, \forall \phi\, \forall \psi\, \big(\mathrm{Sat}(\ulcorner \phi \wedge \psi \urcorner, a) \leftrightarrow (\mathrm{Sat}(\ulcorner \phi \urcorner, a) \wedge \mathrm{Sat}(\ulcorner \psi \urcorner, a))\big)$
- T4 $\forall a\, \forall v \forall \phi\, \big(\mathrm{Sat}(\ulcorner \forall v\, \phi \urcorner, a) \leftrightarrow \forall b\, (\text{‘}b \text{ is } v\text{-variant of } a\text{’} \rightarrow (\mathrm{Sat}(\ulcorner \phi \urcorner, b)\big)$

The theory is equiconsistent with a Tarskian theory of truth, that is, with the theory with the same axioms but the quantifiers $\forall\phi$ and $\forall\psi$ restricted to formulae without Sat. As the theory proves the consistency of Zermelo Fraenkel, it cannot be proved to be outright consistent in ZF, but of course a weakly inaccessible cardinal would suffice. The corresponding theory over arithmetic is investigated under the name FSN in (Halbach 2014, p. 145).

Optional Axioms

Although the theory sketched above is sufficient to prove the results in this paper, the theory is still very weak in the sense that it does not prove any interesting theorems with iterations of the Sat predicate. A theory in the style of the Friedman–Sheard theory is an option (see Halbach 2014). However, for various reasons, I prefer a different approach. Basically I would like to add T1 for Sat as R as long as the formulae to which Sat is applied are *grounded*. This is expressed by the following axiom:

T6 $\forall a\, \forall \phi \forall v\, \Big(\mathsf{G}(\ulcorner \phi \urcorner, a(v)) \rightarrow \big(\mathrm{Sat}(\ulcorner \mathrm{Sat}(\ulcorner \phi \urcorner, v) \urcorner, a) \leftrightarrow \mathrm{Sat}(\ulcorner \phi \urcorner, a(v))\big)\Big)$

Using the truth axioms, one can show the disquotation schema for all grounded sentences.

It remains to specify the axioms for groundedness. I assume that the predicate symbol G is contained as a primitive symbol in the base language, but without any specific axioms (other than in logical axioms and schemata such as replacement); but there is a T1-axiom for G. with the axioms below. All quantifiers (also in the axioms above) are understood to range over all formulae in the language with Sat and G.

- D1 $\forall a\, \forall v\, \forall w\, \big(\mathsf{G}(\ulcorner Rvw \urcorner, a)\big)$
  and similarly for all predicate symbols including Sat, and G

D2 $\forall a \forall \phi \forall v \left(\mathsf{G}(\ulcorner P(\ulcorner \phi \urcorner, v) \urcorner, a) \leftrightarrow \mathsf{G}(\ulcorner \phi \urcorner, a(v)) \right)$, where P is G or Sat

D3 $\forall a \forall \phi \, (\mathsf{G}(\ulcorner \neg \phi \urcorner, a) \leftrightarrow \mathsf{G}(\ulcorner \phi \urcorner, a))$

D4 $\forall a \forall \phi \forall \psi \left(\mathsf{G}(\ulcorner \phi \wedge \psi \urcorner, a) \leftrightarrow (\mathsf{G}(\ulcorner \phi \urcorner, a) \wedge \mathsf{G}(\ulcorner \psi \urcorner, a)) \right)$

D5 $\forall a \forall v \forall \phi \left(\mathsf{G}(\ulcorner \forall v \, \phi \urcorner, a) \leftrightarrow \forall b \, (\text{‘}b \text{ is } v\text{-variant of } a\text{’} \to (\mathsf{G}(\ulcorner \phi \urcorner, b) \right)$

A full analysis of the corresponding theory over arithmetic is given by Fujimoto and Halbach (2018). They show that the theory is ω-consistent. The model construction can be lifted to to set theory if some weak large cardinal axiom is assumed (or even much less).

APPENDIX 2: PROOF OF SUBSTITUTIONAL COMPLETENESS

In this appendix I sketch a proof of the following claim:

> *Substitutional Completeness* · In Ω the following statement is provable: For any model M there is a substitution function I and a variable assignment a such that the following equivalence holds for all sentences ϕ :
>
> $$M \vDash \phi \text{ iff } \mathrm{Sat}(I(\phi), a)$$

I assume the language contains only predicate symbols, but no function symbols. Let a model M be given. First I define the substitution function. The function I replaces an atomic formula Pxy where P is a binary predicate symbols with $\langle x, y \rangle \in v$ where v is a variable associated with the predicate symbol P. Other atomic formulae are dealt with in the same way. With each predicate symbol a new variable is associated. The substitution function relativizes every occurrence of a quantifier $\forall x$ with $x \in z$, where z is again a new variable. As was mentioned above in the definition of substitution functions, variables may have to be renamed to ensure that the variables are really ‘new’. I skip the tedious details.

The definition of the substitution function does not depend on the model M. It remains to define the variable assignment. a assigns the domain of the model M to the ‘relativizing’ variable z. The effect is that quantifiers in $I(\phi)$ now range only over the domain of M under the variable assignment thus defined. If v is the variable associated with the predicate symbol P, a assigns the extension of P in M to the variable v. An induction on the complexity of ϕ establishes the claim.[30]

[30] I am grateful to Walter Dean, Kentaro Fujimoto, Øystein Linnebo, Beau Mount, Carlo Nicolai, Lavinia Picollo, Ian Rumfitt, Thomas Schindler, Göran Sundholm, Albert Visser, Paul Weingartner, Philip Welch, Timothy Williamson, and anonymous referees for discussions about the themes of this paper and for comments on drafts. I also thank the audiences of my talks about this material for their valuable hints and questions.

Aho, Tuomo and Mikko Yrjönsuuri (2009), Late medieval logic, *in* L.Haaparanta, ed., 'The Development of Modern Logic', Oxford University Press, Oxford, pp. 11–77.

Asay, Jamin (2013), *The Primitivist Theory of Truth*, Cambridge University Press.

Asmus, Conrad and Greg Restall (2012), History of logical consequence, *in* F. P.Dov M. Gabbay and J.Woods, eds, 'Logic: A History of its Central Concepts', Vol. 11 of *The Handbook of the History of Logic*, Elsevier, Amsterdam, pp. 11–62.

Bays, Timothy (2001), 'On Tarski on models', *Journal of Symbolic Logic* 66, 1701–1726.

Beall, JC and Greg Restall (2006), *Logical Pluralism*, Clarendon Press.

Behmann, Heinrich (1922), 'Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem', *Mathematische Annalen* 86, 163–229.

Davidson, Donald (1984), *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford.

Dutilh Novaes, Catarina (2012), 'Form and Matter in Later Latin Medieval Logic: The Cases of *Suppositio* and *Consequentia*', *Journal of the History of Philosophy* 50, 339–364.

Eder, Günter (2016), 'Boolos and the metamathematics of Quine's definitions of logical truth and consequence', *History and Philosophy of Logic* 32, 170–193.

Etchemendy, John (1990), *The Concept of Logical Consequence*, Harvard University Press, Cambridge, Massachussets.

Feferman, Solomon (1991), 'Reflecting on Incompleteness', *Journal of Symbolic Logic* 56, 1–49.

Field, Hartry (2015), What is logical validity?, *in* C. R.Caret and O. T.Hjortland, eds, 'Foundations of Logical Consequence', Oxford University Press, Oxford, pp. 33–70.

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

Fujimoto, Kentaro and Volker Halbach (2018), 'Classical grounded truth'. draft.

Halbach, Volker (2014), *Axiomatic Theories of Truth*, revised edn, Cambridge University Press, Cambridge. first edition 2011.

Hubien, Hubert (1976), *Iohannis Buridani Tractatus de Consequentiis*, Vol. XVI of *Philosophes Médievaux*, Publications Universitaires, Louvain.

Jeffrey, Richard C. (1992), *Formal Logic: Its Scope and Limits*, third edn, MGraw–Hill.

Kreisel, Georg (1965), Mathematical logic, *in* T. L.Saaty, ed., 'Lectures on Modern Mathematics III', Wiley, New York, pp. 95–195.

Kreisel, Georg (1967*a*), Informal rigour and completeness proofs, *in* I.Lakatos, ed., 'The Philosophy of Mathematics', North Holland, Amsterdam, pp. 138–171.

Kreisel, Georg (1967*b*), Mathematical logic: what has it done for the philosophy of mathematics?, *in* R.Schoenman, ed., 'Bertrand Russell: Philosopher of the Century', George Allen & Unwin, London, pp. 201–272.

Kreisel, Georg (1980), 'Kurt Gödel: 28 April 1906 – 14 January 1978', *Biographical Memoirs of Fellows of the Royal Society* 26, 148–224.

Linnebo, Øystein (2012), 'How to prove soundness', lecture at Oxford University on 25 April.

McGee, Vann (1992), 'Maximal consistent sets of instances of Tarski's schema (T)', *Journal of Philosophical Logic* 21, 235–241.

Quine, Willard Van Orman (1970), *Philosopy of Logic*, Harvard University Press, Cambridge, Massachusetts. second edition 1986.

Read, Stephen (1994), 'Formal and material consequence', *Journal of Philosophical Logic* 23, 247–265.

Read, Stephen (2012), 'The medieval theory of consequence', *Synthese* 187, 899–912.

Reinhardt, William (1986), 'Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth', *Journal of Philosophical Logic* 15, 219–251.

Rusnock, Paul and Mark Burke (2010), 'Etchemendy and Bolzano on logical consequence', *History and Philosophy of Logic* 31, 3–29.

Schneider, Hubert H. (1958), 'Semantics of the predicate calculus with identity and the validity in the empty individual-domain', *Portugaliae Mathematica* 17, 85–96.

Smith, Peter (2011), 'Squeezing arguments', *Analysis* 71, 22–30.

Tarski, Alfred (1936*a*), 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica Commentarii Societatis Philosophicae Polonorum* 1. translated as 'The Concept of Truth in Formalized Languages' in Tarski (1956).

Tarski, Alfred (1936*b*), 'Über den Begriff der logischen Folgerung', *Actes du congrès international de philosophie scientifique* 7, 1–11. translation "On the Concept of Logical Consequence" in (Tarski 1956, 409–20).

Tarski, Alfred (1956), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Clarendon Press, Oxford.

Tarski, Alfred, Andrzej Mostowski and Raphael M. Robinson (1953), *Undecidable Theories*, North Holland, Amsterdam.

Tarski, Alfred and Robert Vaught (1956), 'Arithmetical extensions of relational systems', *Compositio Mathematica* 13, 81–102.

Williamson, Timothy (1999*a*), 'Existence and contingency', *Aristotelian Society* sup. vol. 73, 227–234. Reprinted with printer's errors corrected in Proceedings of the Aristotelian Society 100 (2000): 321–343 (117–139 in unbound version).

Williamson, Timothy (1999*b*), 'A note on truth, satisfaction and the empty domain', *Analysis* 59, 3–8.

Williamson, Timothy (2013), *Modal Logic as Metaphysics*, Oxford University Press.