Formal Notes on the Substitutional Analysis of Logical Consequence

Volker Halbach

5th November 2017

Logical consequence in first-order predicate logic is defined substitutionally in set theory augmented with a primitive satisfaction predicate: An argument is defined to be logically valid iff there is no substitution instance with true premisses and a false conclusion. Substitution instances are permitted to contain parameters. It is shown that each model in the sense of model theory corresponds to a substitution instance together with an interpretation of the parameters. Substitutional and model-theoretic consequence are proved to be extensionally equivalent. Variations of the substitutional definition lead to concepts of logical consequence in free logic and fixed-domain semantics. In contrast to the model-theoretic analysis, the substitutional account of logical consequence features an intended interpretation, preserves non-relativized truth, and follows more closely traditional definitions of logical consequence.

1 The Substitutional Analysis of Logical Consequence

In this paper I provide a formal substitutional account of logical consequence. Substitutional notions of consequence have been discussed at least since the middle ages. In what could be called semantic theories of logical consequence (as opposed to proof-theoretic analyses), logical consequence is defined as truth preservations under all interpretations. In model-theoretic semantics, interpretations are conceived as formal set-theoretic models. However, this is only a very recent understanding of *interpretation*. Traditionally, in order to refute the formal validity of an argument, logicians showed that there is a substitution instance with true premisses but a false conclusion. Such an interpretation is a substitutional counterexample to the argument in question. In traditional logic there were no restrictions on the vocabulary that can be used to devise counterexamples.[1]In the present paper I would like to revive this old-fashioned substitutional understanding of *interpretation* and make the informal substitutional account precise in a mathematical setting for first-order predicate logic.

The substitutional account of logical consequence advanced in this paper contrast with earlier substitutional definitions of logical truth and consequence by Quine (1986) and others. Many of them were based on the Hilbert's and Bernays' (1939) formalized completeness theorem. Their theorem shows that, whenever a formula is not provable, it has a substitutional counterexample in the language of arithmetic. However, such an account is hardly usable as a conceptual analysis of logical validity: A non-trivial theorem is required in order to demonstrate that there are sufficiently many counterexamples in the language of arithmetic to arrive at an extensionally correct definition of logical validity. To most logicians such a substitutional definition looks much less plausible as an analysis of logical validity than the model-theoretic that has become the standard definition from the 1950s. It is obvious that every arithmetical substitutional counterexample corresponds to a model-theoretic counterexample; but not every model-theoretic counterexample corresponds to a arithmetical substitutional counterexample: Set theory provide a plethora of interpretations that is not bound by any cardinality. That the arithmetical substitutional account with its limited stock of counterexamples yields the same set of first-order validities as the model-theoretic account with its rich class of interpretations comes as a substantial insight. However, one would expect from an adequate conceptual analysis of logical validity that it is obviously adequate and that seeing its adequacy does not require an equivalence proof with another definition.

The substitutional analysis of logical validity advanced in this paper is fundamentally different from those based on the Hilbert–Bernays theorem. In particular, it provides a very rich class of counterexamples. In fact, it can be shown that every model-theoretic interpretation corresponds to a substitutional interpretation; but not every substitutional interpretation has a set-theoretic model as counterpart. The 'intended' interpretation is an example of a substitutional interpretation that has no model-theoretic counterpart.

The philosophical ramifications of the substitutional account of logical validity are discussed in another paper (Halbach 2017). Here I list only a few advantages of the substitutional definition as a conceptual analysis of logical validity without going into details:

1. On the substitutional account logical consequence is truth preserving. On the usual model-theoretic account it is only truth preserving relative to a given

[1]Many logicians up to this day do not identify formal validity with logical validity; here I do make this identification without further defending it.

model; it does not preserve truth *simpliciter*.

2. There is no interpretation in the sense of model theory that can be seen as the 'intended' interpretation of a sentence in the language of set theory. An intended interpretation is exactly what is missing in the proof of truth preservation on the model theoretic account. The absence of an intended interpretation has led philosophers to speculate on the indefinite extensibility of the universe and the impossibility of quantifying over absolutely everything. If logical validity is understood substitutionally, the intended interpretation is trivial: It is the homophonic translation that maps every sentence to itself.

3. As I mentioned above, logical consequence has been understood substitutionally at least since Buridan (see Halbach 2017). The model-theoretic definition appeared only in the 1950s, presumably it made its first appearance in (Tarski and Vaught 1956). It would be reassuring to have a proof that the traditional notion of logical validity, appropriately understood, does not differ extensionally from the modern model theoretic understanding. To this end I present such an equivalence result as Theorem 1.

4. The domain of an ordinary set-theoretic model is not allowed to be empty. There are good reasons for this restriction. If truth is defined as truth under all variable assignments, then all sentences are true in the empty model. There are ways around the problem of course; but the empty domain is a nuisance in model theory and excluded for technical reasons, although there does not seem to be a good philosophical reason. It would be nice to have an account of logical validity that naturally accommodates the empty domain. The substitutional definition affords this.

5. It has been argued that a proper treatment of quantifiers as logical constants requires that they should not be interpreted as ranging over all objects under all interpretations. Yet in model theory the domain varies from model to model and the quantifiers do not range over all objects in any model. This is easily achievable on the substitutional approach.

2 SATISFACTION

The formal starting point for my account is the language of set theory. I could develop my account on the background of a much weaker theory. But since I am going to compare it with the model-theoretic definition of logical consequence, which is formulated in set theory, I choose a set-theoretic framework.

The language $\mathcal{L}$ of set theory is a first-order language with the predicate symbol $\in$ as its only nonlogical symbol. For simplicity, I restrict the number of connectives and quantifiers: $\forall$ is the only quantifier, $\neg$ and $\wedge$ are the only connectives. The variables are $v_1, v_2, v_3, \ldots$. Letters such as x and y are used as metavariables for them. The existential quantifier and other connectives are metalinguistic abbreviations in what follows. Further predicate symbols and individual constants could easily be added. Function symbols are more awkward; I return to them below. The expansion of $\mathcal{L}$ by a new binary symbol Sat for satisfaction is called $\mathcal{L}_{\mathrm{Sat}}$.

On the model-theoretic account, the notion of *truth in a model* is defined in set theory. By Tarski's theorem on the undefinability of truth, this means that there cannot be an 'intended' set-theoretic model. Simple, 'absolute' truth that is not relativized to some model cannot be defined. As has been mentioned above, this has lead to some dissatisfaction with the model-theoretic definition.[2] In particular, on the model-theoretic account, it is not obvious why a logically true sentence should be true, or why a valid argument cannot have premisses that are (actually) true and a conclusion that is (actually) false.

The present account strongly deviates in this point form the model-theoretic analysis of logical consequence. The substitutional account presented here relies on a primitive axiomatized semantic notion. It is a notion of satisfaction rather than mere truth. No predicate satisfying these conditions can be defined in the language $\mathcal{L}$ of set theory.

I assume that all expressions of the language $\mathcal{L}_{\mathrm{Sat}}$ have been coded in the finite von Neumann ordinals in some natural way. In what follows, I do not distinguish between expressions and their codes. For satisfaction I employ the following set of axioms. The notation is explained below.

(s1) $\forall v\, \forall w\, \forall \mathfrak{a}\, (\mathrm{Sat}(v \underset{.}{\in} w, \mathfrak{a}) \leftrightarrow \mathfrak{a}(v) \in \mathfrak{a}(w))$

(s2) $\forall v\, \forall w\, \forall \mathfrak{a}\, (\mathrm{Sat}(v \underset{.}{=} w, \mathfrak{a}) \leftrightarrow \mathfrak{a}(v) = \mathfrak{a}(w))$

(s3) $\forall \mathfrak{a}\, \forall \phi\, (\mathrm{Sat}(\underset{.}{\neg} \phi, \mathfrak{a}) \leftrightarrow \neg \mathrm{Sat}(\phi, \mathfrak{a}))$

(s4) $\forall \mathfrak{a}\, \forall \phi\, \forall \psi\, \big(\mathrm{Sat}(\phi \underset{.}{\wedge} \psi, \mathfrak{a}) \leftrightarrow (\mathrm{Sat}(\phi, \mathfrak{a}) \wedge \mathrm{Sat}(\psi, \mathfrak{a}))\big)$

(s5) $\forall \mathfrak{a}\, \forall \phi\, \big(\mathrm{Sat}(\underset{.}{\forall} v \phi, \mathfrak{a}) \leftrightarrow (\forall \mathfrak{b}\, (\mathrm{var}(\mathfrak{b}, \mathfrak{a}, v) \rightarrow \mathrm{Sat}(\phi, \mathfrak{b})))\big)$

The quantifiers $\forall v$ and $\forall w$ range over (codes of) variables, $\forall \mathfrak{a}$ over variable assignments, and $\forall \phi$ and $\forall \phi$ over formulae of the expanded language $\mathcal{L}_{\mathrm{Sat}}$. All these

[2] Kreisel (1965, 1967*a*,*b*) developed his *squeezing argument* to overcome the lack of an intended interpretation. Göran Sundholm made me aware of (Kreisel 1965, p. 116f.) and (Kreisel 1967*b*, p. 253ff.)

restrictions can be expressed in set theory. The symbol $\underset{\cdot}{\in}$ stands for the function that yields, applied to variables v and w the formula $v \in w$. This function is expressible in the language of set theory, although it lacks any function symbols and the function has to expressed using a suitable formula. The other underdotted symbols are to be understood in an analogous way. Finally, the formula $\mathrm{var}(\mathfrak{b}, \mathfrak{a}, v)$ expresses that the variable assignment $\mathfrak{b}$ differs from $\mathfrak{a}$ at most in the variable v.

Adding these axioms to Zermelo–Fraenkel set theory and expanding all axiom schemata of ZF yields the theory S. The theory S of satisfaction looks very similar to a 'Tarskian' theory of satisfaction. The recursive clauses of the definition of satisfaction have been turned into axioms à la Davidson. The theory is type-free in the sense that the quantifiers $\forall \phi$ and $\forall \psi$ range over all sentences of $\mathcal{L}_{\mathrm{Sat}}$, including those containing the satisfaction predicate. However, S lacks any axioms that impose any restrictions on the satisfaction of atomic formulae with Sat. Of course, one has to proceed very carefully: The addition of an axiom analogous to (s1) for Sat will yield an inconsistency. There are various way to avoid the inconsistency. The theories FS and CG would serve the purpose (see Halbach 2014 and Halbach and Fujimoto 2018). Here, however, no such additional strengthening is required.

LEMMA 1 (uniform T-sentences) *For all formulae of $\mathcal{L}$, that is for all formulae without* Sat, $\phi(x_0, \dots, x_n)$ *the following holds:*

$$\mathsf{S} \vdash \forall \mathfrak{a}\, (\mathrm{Sat}(\ulcorner \phi(x_0, \dots, x_n) \urcorner, \mathfrak{a}) \leftrightarrow \phi(\mathfrak{a}(x_0), \dots, \mathfrak{a}(x_n)))$$

The proof is by induction on the length of $\phi(x_0, \dots, x_n)$. Axioms (s1) and (s2) provide the induction base.

3 SUBSTITUTION FUNCTIONS

Substitution functions are functions that uniformly replace the nonlogical vocabulary in all formulae of the language with suitable expressions. The language $\mathcal{L}_{\mathrm{Sat}}$ contains only the binary predicate symbols $\in$ and Sat as nonlogical symbols.

To avoid variable clashes, substitution functions will have to rename some variables. Let formulae $\sigma_{\in}(x, y)$, $\sigma_{\mathrm{Sat}}(x, y)$ and possibly $\delta(x)$ be given and let v_n be the variable with the highest index, that occurs in any of the formulae $\sigma_{\in}(x, y)$, $\sigma_{\mathrm{Sat}}(x, y)$ and possibly $\delta(x)$ and is distinct from the displayed ones, that is, x and y. If z is the k-th variable v_k, then z' is the variable v_{k+n}.

The substitution function I based on the three, or possibly two, formulae ($\sigma_{\in}(x, y)$, $\sigma_{\mathrm{Sat}}(x, y)$ and possibly $\delta(x)$) is then defined on the set of all $\mathcal{L}_{\mathrm{Sat}}$-formulae ϕ in the following way:

$$I(\phi) := \begin{cases} \sigma_{\in}(x', y') & \text{if } \phi \text{ is } x \in y, \\ x' = y' & \text{if } \phi \text{ is } x = y, \\ \sigma_{\mathrm{Sat}}(x', y') & \text{if } \phi \text{ is } \mathrm{Sat}(x, y), \\ \neg I(\psi) & \text{if } \phi \text{ is } \neg\psi, \\ I(\psi) \wedge I(\chi) & \text{if } \phi \text{ is } \psi \wedge \chi, \textit{and} \\ \forall x' \, ([\delta(x') \rightarrow] I(\psi)) & \text{if } \phi \text{ is } \forall x \, \psi \end{cases}$$

In the last line a relativizing formula $\delta(x)$ is present if a formula $\delta(x)$ is specified by the substitution function. If there is no such formula, quantifiers are not relativized. A function is a substitution function if it is a substitution function based on three such formulae.

Renaming the variables ensures that the free variables in the formulae $\sigma_{\in}(x, y)$, $\sigma_{\mathrm{Sat}}(x, y)$ and $\delta(x)$ are not accidentally bound by quantifiers already present in ϕ.

Applying a substitution function to a sentence ϕ of $\mathcal{L}_{\mathrm{Sat}}$ does not necessarily result in a sentence again, because the formulae $\sigma_{\in}$, σ_{Sat}, and δ are allowed to contain free variables. In fact, it will be crucial for some proofs below that free variables are allowed as parameters.

4 THE DEFINITION OF LOGICAL CONSEQUENCE

We define logical consequence as a relation between sets of formulae and formulae of $\mathcal{L}_{\mathrm{Sat}}$. The formulae may contain free variables. If free variables are understood as analogues of personal and demonstrative pronouns, then a restriction to sentences is hard to justify. For instance, the following argument should count as logically valid:

> All men are mortal.
> This is a man.
> Therefore this is mortal.

As long as the reference of 'this' is kept fixed between the second premiss and the conclusion, the argument is logically valid, whether the object that is pointed at has a name or not.

The definition of a substitution function can be carried out in ZF. The formula $\mathrm{SubF}(x)$ expresses in ZF that x is a substitution function. Logical consequence is now defined in S in the following way, if $\mathrm{Form}_{\mathrm{Sat}}$ is the set of $\mathcal{L}_{\mathrm{Sat}}$-formulae:

SUBSTITUTIONAL DEFINITION OF LOGICAL VALIDITY

$$\forall x \, \forall y \Big(x \vDash_S y \; :\leftrightarrow \; x \subseteq \mathrm{Form}_{\mathrm{Sat}} \wedge y \in \mathrm{Form}_{\mathrm{Sat}} \wedge$$
$$\forall \mathfrak{a} \, \forall I \left(\mathrm{SubF}(I) \rightarrow (\forall z \in x \, \mathrm{Sat}(I(z), \mathfrak{a})) \wedge \mathrm{Sat}(I(\ulcorner \exists x \, x = x \urcorner), \mathfrak{a}) \rightarrow \mathrm{Sat}(I(y), \mathfrak{a}) \right) \Big)$$

I write $\vDash_S y$ for $\emptyset \vDash_S y$, which expresses that y is a logical truth.

The definition is a straightforward formal rendering of the informal substitutional definition of logical truth: A formula is logically true iff all its substitution instances are always satisfied.

The formula $\delta(x)$ in the definition corresponds to the domain of a set-theoretic model, as it restricts the range of the quantifier. Since I haven't imposed any restrictions on $\delta(x)$, there are, however, two essential differences: The formula $\delta(x)$ may fail to apply to something, for instance, if $x \neq x$ is chosen as $\delta(x)$. In the canonical semantics of first-order predicate logic, in contrast, the domain must not be empty. Only in order to obtain a match between the model-theoretic and the substitutional definition of logical consequence, I have added the extra assumption $\mathrm{Sat}(I\ulcorner\exists x\, x = x\urcorner), \mathfrak{a})$. In S it can be proved that $I(\ulcorner\exists x\, x = x\urcorner) = \ulcorner\exists x'\big(\delta_I(x') \wedge x' = x'\big)\urcorner$, where $\delta_I(x)$ is the domain formula associated with the substitution function I and x' is the variable that is substituted for x by I.

At the other extreme, $\delta(x)$ may fail to define a set, for instance, if $\delta(x)$ is the formula $x = x$. On the model-theoretic account, in contrast, the domain always has to be a set. Of course, we could impose restrictions on $\delta(x)$ that rule out such formulae.

In the end I do not think that any of the two restrictions should be imposed. The exclusion of the empty and class-sized domains are forced upon the model-theoretic account by technical difficulties. Admitting the empty domain in the definition of models is not an insurmountable problem; there are just no variable assignments over the empty domain and we cannot define truth as satisfaction under all variable assignments, because otherwise all sentences would be true in a model with the empty domain. Logicians have defined semantics in free logic that overcome this problem. The other restriction, in contrast, is indispensable for the model-theoretic approach: If proper classes were admitted as domains, we could no longer define the notion of satisfaction in a model. The inductive definition could not be shown to have a fixed point for arbitrary class-sized models. On the substitutional approach these problems do not arise, and there are reason to keep the definition more liberal.

The reader may suspect that there is a third difference between the substitutional and the model-theoretic account with respect to domains. On the latter approach it is not required that the domain is definable by a formula, whereas it is definable by $\delta(x)$ on the substitutional account. This is not a real difference, though, because $\delta(x)$ may contain parameters, as will be shown below.

5 LINKING THE SUBSTITUTIONAL AND MODEL-THEORETIC CONCEPTIONS OF CONSEQUENCE

On the substitutional analysis defended in this paper, every set-theoretic model corresponds to a substitutional model in a sense defined below. The correspondence between set-theoretic and substitutional models is not just elementary equivalence; that is, not only satisfy both models the same sentences of the language (without any additional parameters); the set-theoretic model and its substitutional counterpart make the same formulae true for arbitrary variable assignments. Consequently, for every set-theoretic model there is a substitutional model; but not *vice versa*. In this sense there are 'more' substitutional models than set-theoretic models, although the number of set-theoretic and substitutional models is proper class-size.

A substitutional model is a pair $(I, \mathfrak{a})$ of a substitutional interpretation and a variable assignment. Now for each set-theoretic models $\mathcal{M}$ a corresponding substitutional model $(\mathcal{I}, \mathfrak{a}_{\mathcal{M}})$ is defined. The substitutional interpretation $\mathcal{I}$ is the same for all set-theoretic model $\mathcal{M}$. Only the variable assignment $\mathfrak{a}_{\mathcal{M}}$ depends on $\mathcal{M}$.

The substitutional interpretation $\mathcal{I}$ is based on $\langle x, y\rangle \in v_2$ as $\sigma_{\in}(x, y)$, $\langle x, y\rangle \in v_3$ as $\sigma_{\text{Sat}}(x, y)$ and $x \in v_1$ as $\delta(x)$. As above v_1, v_2, and v_3 are the first three variables. For any variable v_n, v'_n is the variable v_{n+3}. So no variable v'_n can be any of the three variables v_1, v_2, and v_3 . The substitution function $\mathcal{I}$ based on these three formulae looks as follows:

$$\mathcal{I}(\phi) := \begin{cases} \langle x', y'\rangle \in v_2 & \text{if } \phi \text{ is } x \in y, \\ x' = y' & \text{if } \phi \text{ is } x = y, \\ \langle x', y'\rangle \in v_3 & \text{if } \phi \text{ is } \text{Sat}(x, y), \\ \neg\mathcal{I}(\psi) & \text{if } \phi \text{ is } \neg\psi, \\ \mathcal{I}(\psi) \wedge \mathcal{I}(\chi) & \text{if } \phi \text{ is } \psi \wedge \chi, \textit{and} \\ \forall x' \, (x' \in v_1 \rightarrow \mathcal{I}(\psi)) & \text{if } \phi \text{ is } \forall x \, \psi \end{cases}$$

Whether a formula ϕ contains Sat or not, $\mathcal{I}(\phi)$ is always a formula of $\mathcal{L}$, that is, a formula without Sat.

A model for $\mathcal{L}_{\text{Sat}}$ in the usual set-theoretic sense of model theory is a triple (D, E, S): D, the domain, is some set; E, the extension of $\in$, is a set of pairs over D; and S, the extension of Sat, is a set of pairs over D. I write $(D, E, S) \vDash \phi[\mathfrak{a}]$ for the formula $\mathcal{L}$ of set theory expressing that the formula ϕ holds in the model (D, E, S) under the variable assignment $\mathfrak{a}$.

Given a variable assignment $\mathfrak{a}$ and a model $\mathcal{M} := (D, E, S)$, the variable assignment

$\mathfrak{a}_{\mathcal{M}}$ is define in set theory in the following way:

$$\begin{aligned}\mathfrak{a}_{\mathcal{M}}(v_1) &\coloneqq D\\ \mathfrak{a}_{\mathcal{M}}(v_2) &\coloneqq E\\ \mathfrak{a}_{\mathcal{M}}(v_3) &\coloneqq S\\ \mathfrak{a}_{\mathcal{M}}(v_{n+3}) &\coloneqq \mathfrak{a}(v_n)\end{aligned}$$

For every set-theoretic model $\mathcal{M}$ there is an equivalent substitutional model $(\mathcal{I}, \mathfrak{a}_{\mathcal{M}})$ in the following sense:

LEMMA 2 $\mathsf{S} \vdash \forall \phi\, \forall \mathfrak{a} \big((D, E, S) \vDash \phi[\mathfrak{a}] \leftrightarrow \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a}_{\mathcal{M}})\big)$

As before, the quantifiers $\forall \phi$ ranges here over all formulae of $\mathcal{L}_{\mathrm{Sat}}$. Lemma 1 and the fact that $\mathcal{I}(\phi)$ is always in $\mathcal{L}$ are used in the proof by induction on the length of ϕ.

In a nutshell, Lemma 2 show that for every model in the usual sense of model theory there is a substitutional model satisfying exactly the same formulae under any given variable assignment (modulo renaming of variables). The variable assignment $\mathfrak{a}_{\mathcal{M}}$ plays a dual role: On the one hand it supplies the values for the free variables v_1, v_2, and v_3 in $\sigma_{\in}$, σ_{Sat}, and δ, on the other it contains the original variable assignment $\mathfrak{a}$.

Lemma 2 immediately establishes that substitutional validity implies ordinary model-theoretic validity: If there is a model $\mathcal{M}$ that makes all sentences in Γ true and ϕ false, then $(\mathcal{I}, \mathfrak{a}_{\mathcal{M}})$ refutes the substitutional validity $\Gamma \vDash_S \phi$, where all elements of Γ and ϕ are $\mathcal{L}_{\mathrm{Sat}}$-sentences:

LEMMA 3 $\Gamma \vDash_S \phi$ *implies* $\Gamma \vDash \phi$

The converse is also true: Set-theoretic validity implies substitutional validity. There are different ways to show this. The differences between the proof strategies matter when generalizations of the account here are considered, for instance, when a theory much weaker than full set theory are considered or the substitutional approach is extended to languages with higher-order or generalized quantifiers.

Kreisel's (1967*a*) squeezing argument shows that intuitive validity extensionally coincides with usual model-theoretic validity. Now substitutional validity nicely slots into the place of intuitive validity in the squeezing argument. First, it is shown that some chosen calculus, say some Hilbert-style calculus, is sound with respect to substitutional validity:

LEMMA 4 $\Gamma \vdash \phi$ *implies* $\Gamma \vDash_S \phi$.

The proof is by induction on the length of proofs and makes use of the axioms for $\mathrm{Sat}(x, y)$ and relies on induction on a formula with the satisfaction predicate.

To complete the squeezing argument, the usual completeness theorem is invoked. That is, $\Gamma \vDash \phi$ implies $\Gamma \vdash \phi$. Combining this with Lemmata 4 and 3 establishes that $\Gamma \vDash \phi$, $\Gamma \vdash \phi$, and $\Gamma \vDash_S \phi$ are all equivalent. In particular, we have the equivalence of model-theoretic and substitutional validity:

THEOREM 1 (equivalence of model-theoretic and substitutional consequence)

$$\Gamma \vDash_S \phi \textit{ iff } \Gamma \vDash \phi$$

Here ϕ *and all elements of* Γ *can be arbitrary* $\mathcal{L}_{\mathrm{Sat}}$*-sentences.*

The detour via provability and Gödel completeness for the right-to-left direction can be avoided. Let a substitutional model $(I, \mathfrak{a})$ with a relativizing formula $\delta(x)$ be given. If the domain formula $\delta(x)$ of the substitutional model $(I, \mathfrak{a})$ defines a set, a corresponding set-theoretic model can easily be defined. If the domain is a proper class, one can show a lemma similar to the Löwenheim–Skolem downwards theorem which yields a substitutional set-sized 'elementary submodel' of $(I, \mathfrak{a})$. For this argument no class theory is assumed and talk about proper classes is just shorthand.

The reader who is worried about the use of a type-free theory of satisfaction may wonder to what extent the proof of Theorem 1 depends on the use of the untyped axioms for Sat. Lemma 2 and therefore also Lemma 3 could have been proved in a typed theory of satisfaction, that is, we could have restricted the quantifiers $\forall \phi$ and $\forall \psi$ in Axioms (S3)–(S5) to sentences of $\mathcal{L}$. This is because $\mathcal{I}(\phi)$ is always a sentence of $\mathcal{L}$ even if ϕ does contain the satisfaction predicate. In contrast, the soundness Lemma 4, which allows one to pass from $\Gamma \vdash \phi$ to $\Gamma \vDash_S \phi$, requires the type-free axioms. If the quantifiers in in Axioms (S3)–(S5) were restricted to sentences of $\mathcal{L}$, Lemma 4 could be proved only for sentences of $\mathcal{L}$, but not for sentences containing Sat. This means that the equivalence

$$\Gamma \vDash_S \phi \text{ iff } \Gamma \vDash \phi$$

could be proved in a typed theory of satisfaction, as long as ϕ and all elements of Γ are in $\mathcal{L}$. Basically, the notion $\vDash_S$ of substitutional consequence could be applied meaningfully only to Sat-free sentences in a type-free setting. Such an approach would resemble Tarski's (1936) old, pre-model-theoretic definition of logical consequence. In contrast to the modern model-theoretic definition of logical consequence, the notion of logical validity would be restricted to an object language and not be universal. I retain universality by using a type-free theory of satisfaction.

6 THE INTENDED INTERPRETATION

Philosophers have agonized about the elusive intended model V of set theory. If one is serious about set theory as one's overall theory and does not assume another class theory on top of it, then there is no intended model. On the substitutional account, in contrast, the intended substitutional interpretation is easily defined: The intended substitutional interpretation just does not substitute anything and leaves all formulae alone: On the intended interpretation formulae are understood at face value and nothing is reinterpreted. As there is no relativizing formula $\delta(x)$ and $\sigma_{\in}$, σ_{Sat} are $\in$ and Sat respectively, no renaming of variables is required. On the substitutional account, the intended interpretation is the identity function on the set of formulae; it is the simplest interpretation, as one would expect from the intended interpretation.

It is obvious that substitutional logical truth implies truth and substitutional consequence preserves truth:

LEMMA 5 *1.* $\mathsf{S} \vdash \forall \phi \in \mathrm{Sent}_{\mathrm{Sat}} \big(\vDash_S \phi \rightarrow \forall \mathfrak{a}\, \mathrm{Sat}(\phi, \mathfrak{a}) \big)$

2. $\mathsf{S} \vdash \forall \Gamma \subseteq \mathrm{Sent}_{\mathrm{Sat}}\ \forall \phi \in \mathrm{Sent}_{\mathrm{Sat}} \big(\Gamma \vDash_S \phi \rightarrow \forall \mathfrak{a} \left(\forall \gamma \in \Gamma\, \mathrm{Sat}(\gamma, \mathfrak{a}) \rightarrow \mathrm{Sat}(\phi, \mathfrak{a}) \right) \big)$

For Kreisel (1967*a*) the absence of an intended interpretation from the model-theoretic definition of logical validity was the main reason for distinguishing it from the 'intuitive' concept of logical validity. The substitutional analysis does not suffer from the same problem, and I would like to advocate it as an explication of the the intuitive concept.

7 VARYING THE BASE THEORY.

I have chosen set theory as the basis for S in order compare it with its main rival, the model-theoretic conception of logical validity. In particular, the choice of the base theory makes it possible to show that for any model in the set-theoretic sense there is a corresponding substitutional model. However, it is also easily possible to formulate the substitutional definition in other theories; this is often less straightforward for the model-theoretic definition.

For instance, instead of Zermelo-Fraenkel set theory, its weakened variant without the axiom of infinity can be used. The existence of variable assignments as arbitrary functions, conceived as sets of ordered pairs, from the set of variables or their indices can no longer be proved. However, in the absence of the axiom of infinity, variable assignment can be defined as a function with a subset of the set of variables or their indices as its domain. One can think of these function still as assigning values to every variable by setting $\mathfrak{a}(x)$ equal to some arbitrary object, say $\emptyset$, if the variable x is not in the domain of the finite function. This facilitates the formulation of the Sat axioms. Otherwise, if $\mathfrak{a}(x)$ is undefined fro some$x$, one would have to add conditions in the Sat axioms requiring that the variable assignment assigns values to all variables occurring freely in the formula in question. This will make notation more clumsy.

If some arithmetical system is used as base theory, similar tricks can be applied. Otherwise the substitutional definition of logical validity can be reformulated in the obvious way. Perhaps even more appropriate would be a theory of syntax as base theory.

So far I have considered a language of set theory with only one nonlogical symbol, the membership symbol $\in$. Other predicate symbols can be treated analogously. Constants, can also be accommodated in the definition of logical consequence. First, the definition of a substitution function is tweaked so that constants can be substituted with arbitrary terms, including variables (avoiding variable clashes). Then the definition of logical consequence is modified to ensure that c always denoted an object in the extension of the relativizing formula in the relevant substitutional models. If there is an individual constant c in the language, for instance, an additional clause is added expressing that the object denoted by c is in the extension of the relativizing formula:

$$\begin{aligned}
&\forall x\,\forall y\Big(x \vDash_S y \;:\leftrightarrow\; x \subseteq \mathrm{Form}_{\mathrm{Sat}} \land y \in \mathrm{Form}_{\mathrm{Sat}} \land \\
&\quad \forall \mathfrak{a}\,\forall I\,\big(\mathrm{SubF}(I) \rightarrow (\forall z \in x\,\mathrm{Sat}(I(z),\mathfrak{a})) \land \\
&\qquad \mathrm{Sat}(I(\ulcorner \exists x\, x = x\urcorner),\mathfrak{a}) \land \mathrm{Sat}(I(\ulcorner \exists x\, x = c\urcorner),\mathfrak{a}) \rightarrow \mathrm{Sat}(I(y),\mathfrak{a})\big)\Big)
\end{aligned} \tag{1}$$

The condition $\mathrm{Sat}(I(\ulcorner \exists x\, x = c\urcorner),\mathfrak{a})$ is equivalent to $\mathrm{Sat}(\ulcorner \delta_I(c)\urcorner,\mathfrak{a})$. Additional constants can be dealt with in the same way.

Function symbols with higher arity are less easy to deal with: In a nutshell, function symbols should not only be replaced only with function symbols of the appropriate arity, because there may be functions in the language expressible by some formula (with parameters) but not with a functional expression. The methods used to define relative interpretations for languages with function symbols can be employed.

Finally it may be asked how strong a theory needs to be for defining a substitutional notion of logical consequence that is provably equivalent to the usual model-theoretic notion. The problem is that the usual model-theoretic notion of consequence will not be definable in weaker theories, at least not in a straightforward way. Here I do not go into these weaker theories. The formalized completeness theorem will become relevant in such situations and the account becomes close to substitutional theories in Quine's (1986) style. At any rate, the substitutional theory is less tied to set theory in its formulation than the model-theoretic definition.

8 FREE LOGIC AND CONSTANT DOMAINS

Some of the complications in the substitutional definition of logical validity arise in connection with the relativizing formula $\delta_I(x)$ of a substitution function I. There are two ways to eliminate at least some of these complications.

First, the definition of a substitution function could be simplified by dropping the relativizing formula completely. This is not just a technical simplification. Relativizing formulae (on the substitutional definition) and domains (on the model-theoretic definition) may be seen as varying interpretations of the quantifiers. A proponent of the simplified definition of a substitution function may argue that quantifiers are logical constants and therefore should not be reinterpreted in any way. On the substitutional approach this is actually very easy, easier than with relativizing formulae. On the model-theoretic account, it is not so straightforward, because we cannot use the set of all objects as our domain; there is no such set. However, one can go higher-order and define logical consequence in a higher-order language. The set of all objects becomes then a higher-order object. Tarski's (1936) original theory of logical consequence was of this kind. More recently, Williamson (2000) defended such an approach.

If relativizing formulae are removed from the definition of substitution function, sentences such as $\exists x \, \exists y \, x \neq y$ become logically valid. More generally, the sentences expressing that there are at least n many objects become logical truths. This phenomenon is well known from Etchemendy's (1999) attack on Tarski's (1936) definition of logical consequence; a related problem was spotted earlier by Hinman et al. (1968) for a fixed finite domain.

As mentioned at the beginning of the section, there are at least two ways to simplify the substitutional definition with respect to the relativizing formulae. On the second, the relativizing formulae are retained, but the restriction that a relativizing formula has to be satisfied by some object is dropped. That is, substitution functions are defined as usual, but the definition of logical consequence is simplified as follows:

$$\forall x \, \forall y \Big(x \vDash_S y \; :\leftrightarrow x \subseteq \mathrm{Form_{Sat}} \wedge y \in \mathrm{Form_{Sat}} \wedge \\ \forall \mathfrak{a} \, \forall I \big(\mathrm{SubF}(I) \rightarrow (\forall z \in x \, \mathrm{Sat}(I(z), \mathfrak{a})) \rightarrow \mathrm{Sat}(I(y), \mathfrak{a}) \big) \Big)$$

Under the model-theoretic definition, this corresponds to individual constants not denoting an object in the domain. Unsurprisingly, this yields free logic. Sentences such as $\exists x \, x = c$ or $\exists x \, x = x$ are no longer logically valid.

This is another advantage of the substitutional account. On the usual model-theoretic definition, the empty domain is excluded because it causes (solvable) problems. If the domain of a set-theoretic model is empty, there are no variable assignments over this domain. Therefore, if truth is defined as satisfaction under all variable assignments, every sentence is true in the model with the empty domain as Schneider (1958) observed. Of course, there are very well-known workarounds, but I assume that the difficulties with the empty domain has lead logicians (and Tarski as the founder of model theory) to exclude the empty domain. But I consider this exclusion merely as a technical quirk of model theory. All problems disappear if the substitutional definition is employed. If logical validity is defined substitutionally, free logic arises naturally; mimicking the technical quirk of the model-theoretic definition and excluding the empty domain causes additional complications.

The languages considered in free logic usually feature individual constants. These constants may fail to denote anything and correspond with singular terms or names in natural language that do not denote anything. The definition of logical consequence with an individual constant c can be simplified by dropping the requirement that c denotes an object by removing the clause $\mathrm{Sat}(I(\ulcorner \exists x\, x = c \urcorner), \mathfrak{a})$. The resulting free logic is positive in the sense that an atomic sentence Pc where P is some predicate symbol can still be satisfied even though c denotes an object that does not satisfy the relativizing formula $\delta_I(x)$ of a given substitution function I. In a negative free logic all such formula would be declared false. To obtain a negative free logic, one can modify the definition of a substitution function by attaching the extra conjunct $\delta_I(c)$ to any translation of an atomic formula containing the constant c.

Generally, many of the moves that can be made in model-theoretic semantics can be mirrored in the substitutional framework advanced here.

BIBLIOGRAPHY

Berka, Karel and Lothar Kreiser (1986), *Logik-Texte: Kommentierte Auswahl zur Geschichte der modernen Logik*, fourth, revised edition edn, Akademie-Verlag, Berlin. First edition 1971, third, enlarged edition 1983.

Etchemendy, John W. (1999), *The Concept of Logical Consequence*, second edition edn, CSLI, Stanford. First edition Cambridge (Mass.): Harvard University Press, 1990.

Halbach, Volker (201?), 'The substitutional theory of logical consequence', *???* ??, ??–??

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

Halbach, Volker (2017), 'The substitutional analysis of logical consequence'. submitted.

Halbach, Volker and Kentaro Fujimoto (2018), 'Classical grounded truth', *??* ??, ??–??

Hilbert, David and Paul Bernays (1939), *Grundlagen der Mathematik*, Vol. II, Springer, Berlin.

Hinman, Peter G., Jaegwon Kim and Stephen P. Stich (1968), 'Logical truth revisited', *Journal of Philosophy* 65, 495–500.

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.

Quine, Willard V. (1986), *Philosophy of Logic*, second edn, Harvard University Press.

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

Tarski, Alfred (1936), 'Über den Begriff der logischen Folgerung', *Actes du congrès international de philosophie scientifique* 7, 1–11. Reprinted (abbreviated) in (Berka and Kreiser 1986, 404–13); also reprinted as "On the Concept of Logical Consequence" in (Tarski 1983, 409–20); page references are given for the translation.

Tarski, Alfred (1983), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, second edition edn, Hackett Publishing, Indianapolis.

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

Williamson, Timothy (2000), 'Existence and contingency', *Proceedings of the Aristotelian Society* 100, 321–343.

Volker Halbach
New College
Oxford OX1 3BN
England
VOLKER.HALBACH@NEW.OX.AC.UK