

Formal Notes on the Substitutional Analysis of Logical Consequence

Volker Halbach

Abstract 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. Variants of this definition of logical consequence are given: Logical validity can be defined with or without identity as a logical constant, and quantifiers can be relativized in substitution instances or not. It is shown that the resulting notions of logical consequence are extensionally equivalent to versions of first-order provability and thus model-theoretic consequence. Every model-theoretic interpretation has a substitutional counterpart, but not vice versa. In particular, in contrast to the model-theoretic account, there is a trivial intended interpretation on the substitutional account, namely the homophonic interpretation that does not substitute anything. Applications to free logic, second-order logic, and theories and languages other than set theory are sketched.

1 The Substitutional Analysis of Logical Consequence

In what could be called semantic theories of logical consequence – as opposed to proof-theoretic analyses –, logical consequence is defined as truth preservation 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 the present paper I attempt to revive the 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 contrasts with earlier substitutional definitions of logical truth and consequence by Quine [20]

2010 Mathematics Subject Classification: Primary 03A05, 03B10

Keywords: logical consequence, logical validity

and others. Some of them were based on the Hilbert and Bernays' [11] formalized completeness theorem, which 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 than the model-theoretic that has become the standard definition from the 1950s.¹ It is obvious that every arithmetical substitutional counterexample corresponds to a model-theoretic counterexample; but not every model-theoretic counterexample corresponds to an arithmetical substitutional counterexample: Set theory provides a plethora of interpretations that are 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. One can restrict the class of counterexamples even further to Δ_2^0 -sets and still prove completeness.² However, one would expect from an adequate conceptual analysis of logical validity that it is *obviously* adequate and that establishing the adequacy of the analysis does not require an ingenious proof.

The substitutional analysis of logical validity advanced in this paper is fundamentally different from those based on arithmetical substitution instances only. In particular, it provides a very rich class of counterexamples. Every model-theoretic interpretation corresponds to a substitutional interpretation; but not every substitutional interpretation has a set-theoretic model as counterpart.

The philosophical ramifications of the substitutional account of logical validity are discussed in [9]. 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 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, because the domain of every model is set-sized and thereby (class-)many sets are excluded from the domain of quantification. If the language included a quantifier expressing that there are proper-class many objects, all claims of the form 'There are class many ...' would be logically false, as McMee [18] observed. 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 lead philosophers to speculate on the indefinite extensibility of the universe or language levels (see, e.g., Rayo [21, sec. 9.6] and Studd [25, sec. 3.6]) and the impossibility of quantifying over absolutely everything (see Williamson [31, sec IV] for a discussion). If logical validity is understood substitutionally, the intended interpretation is trivial: It is the homophonic translation that maps every formula to itself.
3. The notion of substitutional consequence advocated here is universal: It applies to the entire language in which consequence is defined. It shares this advantage with the model-theoretic notion and avoids restricting the notion

of logical consequence to a sublanguage as, for instance, in Tarski's [26] conception, which, in contrast to model theory, permits an intended model and yields truth preservation as an obvious feature of logical consequence.

4. Logical consequence has been understood substitutionally at least since Buridan (see Halbach [9]). The model-theoretic definition appeared only in the 1950s, probably first in Tarski & Vaught [29]. It would be reassuring to have a proof establishing that a more traditional notion of logical validity, appropriately understood, does not differ extensionally from the modern model-theoretic understanding. I prove several results that establish a strong equivalence.
5. Unrestricted quantifiers are logical constants by the usual standards such as Tarski's [28] permutation criterion. Yet their interpretation is varied between set-theoretic models in the sense that the domain of quantification changes from model to model. It has been argued that the proper treatment of unrestricted quantifiers as logical constants requires that their range is kept constant between interpretations (see Williamson [31, p. 327]). This is easily achievable on the substitutional approach.
6. 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.

The model-theoretic definition of consequence for first-order logic is often defended as adequate by appealing to Kreisel's squeezing argument, or the soundness and completeness theorem, more generally.³ Kreisel starts from some informal 'intuitive' notion of logical validity that does not obviously coincide with the model-theoretic notion. Then one argues that some logical calculus is obviously sound with respect to the intuitive notion. Moreover, since every model-theoretic counterexample is also an intuitive counterexample, intuitive validity implies model-theoretic validity. Thus intuitive validity is implied by provability and implies model-theoretic validity; it is therefore 'squeezed' between those two formally defined conceptions of validity. Finally, model-theoretic validity implies provability by the completeness theorem for first-order logic. Consequently, all three notions are extensionally equivalent. I see my account of substitutional consequence as an attempt to provide a formal account of the elusive concept of intuitive validity. It would be very worrying if there were a fundamental reason why intuitive validity could never be captured or explicated by a formal account.

I suspect the main reason why Kreisel and others did not attempt a direct analysis of the intuitive notion of logical validity was a reluctance to endorse the use of a primitive concept of truth or satisfaction. The semantic definition of consequence as truth preservation under all interpretations requires some notion of truth. But for a long time this had to be a reductive, mathematically defined notion; a primitive axiomatized notion was out of the question. For mathematicians there may be good reasons to avoid primitive semantic notions. For the conceptual analysis of logical

validity it can hardly be avoided. Of course, after having shown the equivalence of the direct substitutional analysis with model-theoretic validity, the primitive notion of truth is no longer needed and the model-theoretic definition of logical consequence can be used: The squeezing argument establishes the extensional adequacy of model-theoretic validity and has thus been converted from a necessarily informal argument into a *formally rigorous* proof by replacing Kreisel's informal 'intuitive validity' with the formally rigorous notion of substitutional validity.

Plan of Paper

In the present paper I develop the formal theory of the substitutional definition of logical consequence as I envisage it. In many ways the substitutional account is very flexible and more so than the model-theoretic. This applies especially to the treatment of identity and the quantifiers. I present several variants of the definition of logical consequence. The reader can then choose a version, according to philosophical preferences.

In the following section I introduce the formal theory, an extension of set theory with a primitive satisfaction predicate, in which logical consequence for the language of this theory (and other first-order languages) will be defined. In section 3 the notion of a substitution instance will be precisely defined. The substitutions treat identity as a nonlogical predicate (and thus the identity predicate can be substituted with any formula); quantifiers cannot be relativized by substituting an unrestricted quantifier with a relativized quantifier. Soundness and completeness relative to first-order provability is established. In section 4 I modify the definition of a substitution function by treating identity as a logic constant. Substitution functions permitting relativization of quantifiers are discussed. It is proved that with a suitable notion of substitution function every set-theoretic model corresponds to a substitutional model (suitably defined) in a straightforward way. In the last section I look at some extensions of the substitutional definition of logical consequence and provide a further outlook.

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, a set-theoretic framework serves the purpose of this paper best.

The language \mathcal{L} of set theory is a first-order language with the predicate symbols \in and $=$. \forall is the only quantifier, \neg and \wedge are the only connectives. The variables are $v_0, v_1, v_2, v_3, \dots$. Letters such as x and y are used as metavariables for them. The existential quantifier and other connectives are metalinguistic abbreviations. Further predicate symbols and individual constants could easily be added. Function symbols are more awkward; I return to them later. The expansion of \mathcal{L} by a new binary symbol Sat for satisfaction is called \mathcal{L}_{Sat} .

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

$$(S1) \quad \forall v \forall w \forall a (\text{Sat}(v \in w, a) \leftrightarrow a(v) \in a(w))$$

- (S2) $\forall v \forall w \forall a (\text{Sat}(v \doteq w, a) \leftrightarrow a(v) = a(w))$
 (S3) $\forall a \forall \phi (\text{Sat}(\neg \phi, a) \leftrightarrow \neg \text{Sat}(\phi, a))$
 (S4) $\forall a \forall \phi \forall \psi (\text{Sat}(\phi \wedge \psi, a) \leftrightarrow (\text{Sat}(\phi, a) \wedge \text{Sat}(\psi, a)))$
 (S5) $\forall a \forall \phi (\text{Sat}(\forall v \phi, a) \leftrightarrow (\forall b (\text{var}(b, a, v) \rightarrow \text{Sat}(\phi, b))))$

The quantifiers $\forall v$ and $\forall w$ range over (codes of) variables, $\forall a$ over variable assignments, that is functions with the set of variables as domain, and $\forall \phi$ and $\forall \psi$ over formulae of the expanded language \mathcal{L}_{Sat} . All these restrictions can be expressed in set theory. The symbol \doteq 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 be expressed using a suitable formula. The other underdotted symbols are to be understood in an analogous way. Finally, the formula $\text{var}(b, a, v)$ expresses that the variable assignment b differs from a at most in the variable v .

Adding these axioms to Zermelo–Fraenkel set theory and expanding all axiom schemata of ZF (separation and replacement) to the new language with Sat 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}_{Sat} , including those containing the satisfaction predicate. However, S lacks axioms that impose 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) or (S2) for Sat will yield an inconsistency. There are various ways to avoid the inconsistency. The theories FS and CD would serve the purpose (see Halbach [8] and Halbach & Fujimoto [6]). Here, however, no such additional strengthening is required.

Lemma 2.1 (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:*

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

$a(x)$ stands for the value of x under the assignment a . The proof is by meta-induction on the length of $\phi(x_0, \dots, x_n)$. Axioms (S1) and (S2) provide the induction base.

The next lemma on truth preservation under substitution of identicals, in contrast, is proved by a formal induction in S :

Lemma 2.2 $S \vdash \forall \phi \in \mathcal{L} \forall v \forall w \forall a (a(v) = a(w) \rightarrow (\text{Sat}(\phi, a) \leftrightarrow \text{Sat}(\phi(w/v), a)))$

The expression $\phi(w/v)$ stands for the formal substitution of all occurrences of the variable v that are free for w with w .

The quantifier $\forall \phi \in \mathcal{L}$ expresses quantification over formulae without Sat , and the restriction cannot be relaxed, because there is no axiom for Sat that is analogous to the axioms (S1) and (S2) for $=$ and \in .

3 Absolute Substitution Functions

Absolute or unrelativized substitution functions are functions that uniformly replace the nonlogical vocabulary in all formulae of the language with suitable expressions. For the moment being, I omit the qualification ‘absolute’ and call absolute substitution functions just ‘substitution functions’. Although the language \mathcal{L}_{Sat} contains

only the binary predicate symbols $=$, \in and Sat and thus has very restricted vocabulary for substitution instances, the substitutional definition is sound and complete and therefore extensionally equivalent to the usual model-theoretic definition.⁴

In the present section I treat identity as a nonlogical symbol and therefore in the same way as the other two predicate symbols \in and Sat . Identity as a logical constant will be discussed later.

To avoid variable clashes, substitution functions will have to rename some variables. Let formulae $\sigma_=(x,y)$, $\sigma_\in(x,y)$ and $\sigma_{\text{Sat}}(x,y)$ be given. The three formulae may contain more free variables than those displayed. Let v_n be the variable with the highest index that occurs in $\sigma_=(x,y)$, $\sigma_\in(x,y)$ or $\sigma_{\text{Sat}}(x,y)$ 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+1} .

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

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

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

To specify an absolute substitution function it suffices to specify $\sigma_=(x,y)$, $\sigma_\in(x,y)$ and $\sigma_{\text{Sat}}(x,y)$. Therefore there are countably many substitution functions; and they are all primitive recursive, assuming the coding is reasonable.

Applying a substitution function to a sentence ϕ of \mathcal{L}_{Sat} does not necessarily result in a sentence again, because the formulae $\sigma_=(x,y)$, $\sigma_\in(x,y)$ and $\sigma_{\text{Sat}}(x,y)$ are allowed to contain additional free variables. In fact, it will be crucial for some proofs below that additional free variables are allowed as parameters. In [9] I tried to defend the presence of these free variables by pointing out that in counterexamples to arguments in natural language we can use personal or demonstrative pronouns, if we lack a name for an object. We only need to ascertain that the reference of the pronoun does not change between different occurrences in a substitution instance. As a referee correctly pointed out, the analogy between pronouns and free variables is dubious. Nevertheless, I would maintain that free variables can serve the same purposes as pronouns in natural-language counterexamples.

3.1 The Definition of Logical Consequence The definition of a substitution function can be carried out in ZF. The formula $\text{SubF}(x)$ expresses in ZF that x is a substitution function. Logical consequence is now defined in S in the following way. $\forall \Gamma$ ranges over sets of \mathcal{L}_{Sat} -formulae, $\forall \phi$ over formulae of \mathcal{L}_{Sat} , and $\forall \alpha$ over variable assignments. $\text{Sat}(I(\Gamma), \alpha)$ is short for $\forall \gamma \in \Gamma \text{Sat}(I(\gamma), \alpha)$.

substitutional definition of logical validity (absolute)

$$\forall \Gamma \forall \phi \left(\Gamma \models_S \phi \iff \forall I \forall \alpha \left(\text{SubF}(I) \rightarrow \left(\text{Sat}(I(\Gamma), \alpha) \rightarrow \text{Sat}(I(\phi), \alpha) \right) \right) \right)$$

I write $\models_S \phi$ for $\emptyset \models_S \phi$, which expresses that ϕ is a logically valid.

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.

3.2 Soundness & Completeness I am going to prove that one of the the usual calculi for first-order logic (without any special axioms and rules for identity) is sound and complete with respect to \models_S . Obviously, the proof depends on the chosen calculus. Since the substitutional consequence relation \models_S obtains between sets of formulae and formulae possibly containing free variables, a system is needed that permits free variables in the premiss set and the conclusion. If a Natural Deduction calculus is used, the soundness proof will resemble the informal gloss that is often used to persuade students that the rules of Natural Deduction preserve truth. Undischarged assumptions are added to the premiss set and removed again when they are discharged. An absolute substitution function replaces only atomic formulae with possibly more complex formulae, so that the logical form of the formula is not affected by the substitution.

Within S it is shown by induction on the length of the proof that whenever the premisses are true under a variable assignment, then the conclusion is true under that variable assignment as well. For instance, the soundness of the conjunction introduction rule requires the left-to-right direction of axiom (S4), the rules for negation (S3). For the \forall -introduction rule, one proceeds as follows: If no sentence $I(\gamma)$ with $\gamma \in \Gamma$ contains x and all $I(\gamma)$ are true under \mathfrak{a} and $I(\phi(x))$ is true under \mathfrak{a} , then $I(\phi(x))$ is true under all variable assignments \mathfrak{b} that differ from \mathfrak{a} in x only. Hence $\forall x I(\phi)$, that is $I(\forall x \phi(x))$ is true under \mathfrak{a} . In contrast to the soundness proof for model-theoretic consequence, there is no reference to a domain. This gives soundness of provability in the calculus, denoted by \vdash :

Lemma 3.1 (soundness) *S proves that $\Gamma \vdash \phi$ implies $\Gamma \models_S \phi$.*

A substitution function is parameter-free iff for all ϕ , $I(\phi)$ does not have more free variables than ϕ ; that is, I does not introduce new free variables. In particular, if I is a parameter-free substitution function and ϕ a sentence, then $I(\phi)$ is also a sentence.

Theorem 3.2 (completeness) *There is a parameter-free, absolute substitution function H and an assignment \mathfrak{h} such that*

$$S \vdash \forall \Gamma (Con(\Gamma) \rightarrow Sat(H(\Gamma), \mathfrak{h}))$$

Therefore S proves that $\Gamma \models_S \phi$ implies $\Gamma \vdash \phi$.

As above, $Sat(H(\Gamma), \mathfrak{h})$ is short for $\forall \gamma \in \Gamma Sat(H(\gamma), \mathfrak{h})$. $Con(\Gamma)$ expresses the logical consistency of the set Γ of \mathcal{L}_{Sat} -formulae in the chosen calculus.

The theorem means that parameters are not needed to prove the completeness theorem. Therefore S proves that, if Γ is a consistent set of sentences, then $H(\Gamma)$ is true (satisfied by all assignments).⁵

The rest of this section contains the proof of the completeness theorem by a Henkin construction. First a maximal consistent set Γ_ω sentences with Henkin witnesses is constructed; then a substitutional function rather than a term model is defined from this set.

I assume that the starting set Γ contains only variables with even indices. Thus all variables (whether bound or free) come from the set $\{v_{2n} : n \in \omega\}$ and I call them

even variables. This is no loss of generality: If a set Δ of \mathcal{L}_{Sat} -formulae does contain odd variables, all variables are renamed by doubling the index to obtain a Γ with only even variables and then construct the substitution function H and the variable assignment \mathfrak{h} as below. Finally, one can return to the original set Δ by dividing all variable indices by 2 and adapting the variable assignment by setting $\mathfrak{h}'(v_n) = \mathfrak{h}(v_{2n})$.

The reason for this assumption on variables is that instead of Henkin constants, I add variables as witnesses. That is, for each formula ϕ and even variable x , there is a different odd variable y such that the following formula is added to Γ :

$$\neg\forall x \phi(x) \rightarrow \neg\phi(y)$$

After adding all these formulae, the set is extended to a maximal consistent set in the usual way, treating odd variables as constants. That is, odd constants are not permitted in bound positions. This yields a maximal consistent set Γ_ω of formulae in the usual way.

Now the fork in the road has been reached, with one direction leading to the completeness proof for model-theoretic, and the other to substitutional validity. To prove model-theoretic completeness, I would build a Henkin model \mathcal{H} and a variable assignment \mathfrak{h} over \mathcal{H} such that $\mathcal{H} \models \phi[\mathfrak{h}]$ iff $\phi \in \Gamma_\omega$ for all formulae ϕ by using the set of all (odd and even) variables as domain. Instead of using variables as the elements of the domain, one can also use directly their indices, which will yield exactly set ω of non-negative integers as domain. For substitutional validity the approach is different. Since I deal with absolute substitution functions, there is no corresponding way to use a domain and quantifiers are not restricted in any way by the substitution function. However, one can still use a strategy similar to the construction of the term model with ω as domain by stipulating that all objects not in ω behave in the same way as, say, 0.

To specify the absolute substitution function H , I merely need to fix $\sigma_=(x,y)$, $\sigma_\in(x,y)$, and $\sigma_{\text{Sat}}(x,y)$. The definition is analogous for $=$, \in , and Sat (and any other further predicate symbols that might be added). I use \in as an example.

In set theory the following function can be defined:

$$f(x) := \begin{cases} x & \text{if } x \in \omega, \\ 0 & \text{if } x \notin \omega \end{cases}$$

$\ulcorner v_x \in v_y \urcorner \in \Gamma_\omega$ expresses that the formula $v_i \in v_j$ is in Γ_ω for $i = f(x)$ and $j = f(y)$. This definition ensures that $\ulcorner v_x \in v_y \urcorner \in \Gamma_\omega$ is defined for all arguments, including those that are not finite von Neumann ordinals; such objects are treated as if they were 0.

According to the substitution function H , $\sigma_\in(x,y)$ is the formula $\ulcorner v_x \in v_y \urcorner \in \Gamma_\omega$ and hence the following obtains:

$$H(x \in y) = \sigma_\in(x',y') = (\ulcorner v_{(x')} \in v_{(y')} \urcorner \in \Gamma_\omega)$$

The renaming of variables is awkward. As long as readers trust that variable clashes can be avoided by the usual tricks, they may identify x and x' and ignore the prime symbol.

The variable assignment \mathfrak{h} matching the substitution function H in the completeness theorem is defined by setting

$$\mathfrak{h}(v'_n) := n \text{ for all } n \in \omega. \quad (1)$$

If there is no variable y such that $x = y'$, I set $\mathfrak{h}(x) = 0$. The values of such variables do not matter.

Let $\mathcal{L}_{\text{Sat}}^*$ be the language \mathcal{L}_{Sat} with all formulae that contain bound occurrences of odd variables removed. The main lemma for the proof of the theorem is the following:

Lemma 3.3 $\mathbb{S} \vdash \forall \phi \in \mathcal{L}_{\text{Sat}}^* (\text{Sat}(H(\phi), \mathfrak{h}) \leftrightarrow \phi \in \Gamma_\omega)$

Proof of lemma 3.3 The lemma is proved by induction on the complexity of ϕ . In the case that ϕ is $v_n \in v_k$, I reason in \mathbb{S} as follows:

$$\begin{aligned} \text{Sat}(H(v_n \in v_k), \mathfrak{h}) &\leftrightarrow \text{Sat}(\ulcorner v_{v'_n} \in v_{v'_k} \urcorner \in \Gamma_\omega, \mathfrak{h}) && \text{def. of } H \\ &\leftrightarrow \ulcorner v_{\mathfrak{h}(v'_n)} \in v_{\mathfrak{h}(v'_k)} \urcorner \in \Gamma_\omega && \text{axiom (S1)} \\ &\leftrightarrow \ulcorner v_n \in v_k \urcorner \in \Gamma_\omega && (1) \end{aligned}$$

The connectives are straightforward. The quantifier case is obviously the most interesting. For the formula $\forall v_n \phi(v_n)$ there is an odd Henkin variable y such that $\neg \forall v_n \phi(v_n) \rightarrow \neg \phi(y)$ is in Γ_ω . Let this variable y be fixed in what follows:

$$\begin{aligned} \text{Sat}(H(\forall v_n \phi(v_n)), \mathfrak{h}) &\leftrightarrow \text{Sat}(\forall v'_n H(\phi(v_n)), \mathfrak{h}) && \text{def. of } H \\ &\leftrightarrow \forall \mathfrak{b} (\text{var}(\mathfrak{b}, \mathfrak{h}, v'_n) \rightarrow \text{Sat}(H(\phi(v_n), \mathfrak{b}))) && \text{axiom (S5)} \\ &\rightarrow \text{Sat}(H(\phi(v_n))(y'/v'_n), \mathfrak{h}) && \text{lemma 2.2} \\ &\rightarrow \text{Sat}(H(\phi(y)), \mathfrak{h}) && \text{def. of } H \\ &\rightarrow \phi(y) \in \Gamma_\omega && \text{ind. hyp.} \\ &\rightarrow \neg \phi(y) \notin \Gamma_\omega && \text{consistency} \\ &\rightarrow \neg \forall v_n \phi(v_n) \notin \Gamma_\omega && \text{Henkin property} \\ &\rightarrow \forall v_n \phi(v_n) \in \Gamma_\omega && \text{completeness} \end{aligned}$$

The converse direction can be proved by contraposition:

$$\begin{aligned} \neg \text{Sat}(H(\forall v_n \phi(v_n)), \mathfrak{h}) &\leftrightarrow \exists \mathfrak{b} (\text{var}(\mathfrak{b}, \mathfrak{h}, v'_n) \wedge \neg \text{Sat}(H(\phi(v_n), \mathfrak{b}))) && \text{(S5)} \\ &\rightarrow \exists k \in \omega \neg \text{Sat}(H(\phi(v_n))(v'_k/v'_n), \mathfrak{h}) && \text{lemma 2.2} \end{aligned}$$

The second line follows, because for any \mathfrak{b} and n there is a $k \in \omega$ such that $\mathfrak{b}(v_n) = k = \mathfrak{h}(v'_k)$, if $\mathfrak{b}(v_n) \in \omega$; if $\mathfrak{b}(v_n) \notin \omega$, k can be chosen as 0. The proof is now continued as follows, writing $\phi(v_k)$ for $(\phi(v_n))(v_k/v_n)$:

$$\begin{aligned} &\rightarrow \exists k (\phi(v_k)) \notin \Gamma_\omega && \text{ind. hyp.} \\ &\rightarrow \exists k (\neg \phi(v_k)) \in \Gamma_\omega && \text{completeness} \\ &\rightarrow (\forall v_n \phi(v_n)) \notin \Gamma_\omega && \text{consistency} \end{aligned}$$

□

Lemma 3.3 then yields the Completeness theorem just like in the model-theoretic case.

3.3 Substitutional Validity and Interpretations Lemma 3.1 and theorem 3.2 show that the substitutional definition of logical consequence is extensionally equivalent to the model-theoretic definition; but mere extensional adequacy is not sufficient for a proper conceptual analysis. Truth preservation in all Δ_2^0 -models or in all models of cardinality \aleph_{85} is also extensionally adequate. Therefore theorem 3.2 fails to

show that \models_S is a proper, intensionally adequate analysis of logical validity. Of course, no formal proof can be given that \models_S is a fully adequate analysis. But at least I can show that the substitutional notion of validity does not fall back behind the model-theoretic in the sense that very model-theoretic counterexample is also a substitutional counterexample.

As remarked above, there are only countably many substitution functions. If only parameter-free substitution functions are considered, then model theory offers many more ways of interpreting first-order sentences. Of course, parameter-free functions suffice for proving completeness; but it is one of the standard objections to substitutional conceptions of validity that the ways of interpreting sentences are so much more limited than on the model-theoretic approach. By admitting parameters in substitution functions, the situation is reversed: It can be shown that every model corresponds to a substitution function plus a variable assignment to the parameters; but conversely not every pair of substitution function and variable assignment correspond to a model in the sense of model theory. In the case of absolute substitution functions the following can be shown.

A model for \mathcal{L}_{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 $(D, E, S) \models \phi[\mathfrak{a}]$ for the \mathcal{L} -formula of set theory expressing that the formula ϕ holds in the model (D, E, S) under the variable assignment \mathfrak{a} over D .

Theorem 3.4 *There is an absolute substitutional function I such that for every set-theoretic model $\mathcal{M} := (D, E, S)$ the following holds:*

$$\text{S} \vdash \forall \phi \forall \mathfrak{a} \lambda D ((D, E, S) \models \phi[\mathfrak{a}] \leftrightarrow \text{Sat}(\mathcal{I}(\phi), \mathfrak{a}^I_{\mathcal{M}}))$$

As before, the quantifier $\forall \phi$ ranges over all formulae of \mathcal{L}_{Sat} . $\mathfrak{a} \lambda D$ ranges over all variable assignments over D , $\mathfrak{a}^I_{\mathcal{M}}$ is obtained from the variable assignment \mathfrak{a} by adjusting it according to the renaming of variables: The substitution function I renames variables by mapping x to x' . The variable assignment $\mathfrak{a}^I_{\mathcal{M}}$ satisfies $\mathfrak{a}^I_{\mathcal{M}}(x') = \mathfrak{a}(x)$. Moreover it assigns D , E , and S to the parameters of I .

The proof of the theorem is a simplified version of that of theorem 4.1 below.

All model-theoretic interpretations have a substitutional counterpart, but not conversely. From the model-theoretic perspective, the substitutional account admits ‘proper class-sized domains’. The homophonic or ‘intended’ interpretation is such an interpretation. This claim will be made precise below.

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 σ_{\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 3.5**
1. $S \vdash \forall \phi \in \text{Sent}_{\text{Sat}} (\models_S \phi \rightarrow \forall \alpha \text{Sat}(\phi, \alpha))$
 2. $S \vdash \forall \Gamma \subseteq \text{Sent}_{\text{Sat}} \forall \phi \in \text{Sent}_{\text{Sat}} (\Gamma \models_S \phi \rightarrow \forall \alpha (\forall \gamma \in \Gamma \text{Sat}(\gamma, \alpha) \rightarrow \text{Sat}(\phi, \alpha)))$

The intended interpretation is obviously parameter-free. Therefore the variable assignment is superfluous in the lemma if ϕ and all elements of Γ are sentences. Analogous claims hold for absolute substitution functions.

4 Identity and Relativized Substitution Functions

In the substitutional, the model-theoretic, and Tarski's [26] theory, logical consequence is defined as truth preservation under all interpretations of the nonlogical vocabulary. I call all these theories *semantic*.⁶ When defining logical validity in a semantic way with identity as a logical constant, one faces a difficult decision: Either one treats quantifiers differently from other logical constants and permits their reinterpretation by varying their range, *or* one treats quantifiers truly as constants with no permissible reinterpretation, but also accepts that $\exists x \exists y x \neq y$ and generally all statements that there are n many objects (for each n) are logical truths.⁷ It is not clear that there is a correct choice here. Logical consequence may here fan out into several conceptions. I would strongly resist only the view that $\exists x \exists y x \neq y$ is not logically valid *because* it is metaphysically possible that there is only one or even no object at all. That is, if one rejects the logical validity of $\exists x \exists y x \neq y$, then this must not be because it possibly metaphysically false, but for some other reason. First, I maintain that it is not metaphysically possible that $\exists x \exists y x \neq y$ fails; and secondly I would defend the big progress of medieval logicians to purge the notion logical consequence from dark notions of necessity (even if they usually flinched from contradicting Aristotle by identifying formal with logical validity). What I am after is formal validity untainted from other modal notions.

The dominant view is, of course, that sentences such as $\exists x \exists y x \neq y$ are not logical truths and that the range of quantifiers can be varied among interpretations. On the model-theoretic account, there is no alternative to restricting the domain of quantification to some set: The domain of quantification in a model *has* to be a set; otherwise satisfaction in that model is not definable within set theory. On the substitutional account it is possible to leave quantifiers unrelativized under interpretations, as I have done with the absolute substitution functions, while identity was treated as a nonlogical expression, thereby avoiding the logical validity of $\exists x \exists y x \neq y$.

Now with identity as logical constant, one can either stick to absolute substitution functions and accept that $\exists x \exists y x \neq y$ is a logical truth or relativize quantifiers. Here I prefer to avoid taking a stance and show that both choices are available on the substitutional account by slight modifications of the definition of an absolute substitution function.

The system S contains already the Sat-axiom (S2) for identity. This axiom is analogous to the axioms for nonlogical predicates such as \in and Sat . Together with the other axioms for Sat , it proves substitution of identicals – but only for formulae not containing Sat :

$$S \vdash \forall \phi \in \mathcal{L} \forall v \forall w \forall \alpha (\alpha(v) = \alpha(w) \rightarrow (\text{Sat}(\phi, \alpha) \leftrightarrow \text{Sat}(\phi(w/v), \alpha)))$$

The axioms for the other logical constants, that is the connectives and the quantifier, have been stated in a type-free way. For instance, a conjunction is true iff both

conjuncts are true, irrespective of whether they contain the Sat-predicate or not. Consequently, it can be shown that substituting equivalent formulae does not affect the truth or falsity of a sentence. With the axioms mentioned so far, only a typed version of the principle of substitution of identicals is provable. However, the Sat-axioms for identity should follow the policy of treating logical constants as untyped. Therefore I postulate the following axiom, expressing the the principle of substitution of identicals in a type-free way for all formulae ϕ , including those with Sat:⁸

$$(S=) \quad \forall \phi \forall v \forall w \forall a (a(v) = a(w) \rightarrow (\text{Sat}(\phi, a) \leftrightarrow \text{Sat}(\phi(w/v), a)))$$

It would have been sufficient to postulate this for atomic formulae of the form $\text{Sat}(x, y)$ instead of ϕ ; the generalized form above follows then by formal induction.

Several other axiomatic theories of satisfaction and truth fail to prove this axiom, unless it or something similar is added explicitly. (S=) is required for the soundness proofs below, that is, to prove that all substitution instances of the satisfaction axioms are true (satisfied by all assignments).

4.1 Identity and Absolute Interpretation Functions An absolute substitution function with identity as logical constant is defined just like the substitution functions above, except that I also demand that $I(x = y)$ is always the formula $x' = y'$ for every such substitution function I . A formula $\sigma_{=} (x, y)$ is consequently no longer needed.

Since quantifiers remain unmodified by such substitution functions, $\exists x \exists y x \neq y$ becomes logically valid, as remarked above. For the adequacy theorem the notion of provability has to be changed accordingly. Besides adding the usual rules and/or axioms for identity, all sentences stating that there are at least n -many pairwise distinct objects are added as logical axioms. For this modified notion of the logical calculus soundness and completeness can be proved along the lines of the proof above. Rayo & Williamson [22] proved a similar completeness theorem for a consequence relations defined in the style of Tarski [26].

4.2 Relativized Interpretation Functions To obtain the usual notion of validity in first-order logic with identity, while avoiding $\exists x \exists y x \neq y$ as a logical truth, I modify the definition of a substitution function in a way that can simulate the domains of the model-theoretic definition. A relativized substitution function I is based on two or three formulae ($\sigma_{\in} (x, y)$, $\sigma_{\text{Sat}} (x, y)$) and possibly on a formula $\delta(x)$. The renaming function mapping x to x' has to be adapted to the third formula. A substitution function I based on those two or three formulae is then defined in the following way:

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

In the last line the square brackets around $\delta(x)$ mean that the (uniform) relativization of the quantifiers is optional. Hence absolute substitution functions are a special case of the relativized functions. We could also insist on a relativizing formula for all substitution functions. By choosing $x = x$ as $\delta(x)$ absolute substitution functions could be emulated. However, I prefer to retain the homophonic interpretation without relativization of the quantifier.

Using the relativized substitution functions, I define validity in such a way that it extensionally coincides with model-theoretic validity. Again identity causes some trouble.

The formula $\text{SubFr}(x)$ expresses in ZF that x is a relativized substitution function. Logical consequence is now defined in S in the following way:

substitutional definition of logical validity (relativized)

$$\forall \Gamma \forall \phi \left(\Gamma \models_S \phi : \leftrightarrow \right. \\ \left. \forall I \forall \alpha \left(\text{SubFr}(I) \rightarrow (\text{Sat}(I(\Gamma), \alpha)) \wedge \text{Sat}(I(\ulcorner \exists x x = x \urcorner), \alpha) \rightarrow \text{Sat}(I(\phi), \alpha) \right) \right)$$

The formula $\delta(x)$ in a relativized substitution function corresponds to the domain of a set-theoretic model, as it restricts the range of the quantifier. Since I have not imposed any restrictions on $\delta(x)$, there are two essential differences to the domain in the model-theoretic definition: The formula $\delta(x)$ may fail to apply to anything, 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 $\text{Sat}(I(\ulcorner \exists x x = x \urcorner), \alpha)$. In S it can be proved that $I(\ulcorner \exists x x = x \urcorner) = \ulcorner \exists x' (\delta_I(x') \wedge x' = x') \urcorner$, 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 restriction to non-empty sets as domains is 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: The quantifiers have to be interpreted as ranging over some set, because the definition of satisfaction requires that models have a (set-sized) domain. On the substitutional approach, in contrast, the relativizing formula may be completely absent or a formula applying to objects that do not form a set.

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 and, of course, every set is definable with set parameters; this will be used in the following section.

4.3 Substitutional Counterparts of Set-Theoretic Models For the proof of the completeness of \models_S relative to some calculus for predicate logic with identity, I could give another Henkin-style proof. However, it is easy to show that each set-theoretic

model corresponds to a substitutional model, that is, a relativized substitution function plus a variable assignment. Completeness follows then from completeness theorem for the model-theoretic definition of consequence. The proof that every set-theoretic model has a substitutional counterpart is also required to demonstrate that nothing is lost on the substitutional account compared to the model-theoretic. It shows that the present version of substitutional validity can overcome the worry that there are ‘fewer’ substitutional than set-theoretic interpretations. Whatever constitutes a possible counterexample on the model-theoretic account is also – modulo some trivial rewriting – a substitutional counterexample, while the converse does not hold. This will be discussed below.

I will now show how a set-theoretic model can be converted into a substitutional model, where a substitutional model is a pair $\langle I, \alpha \rangle$ of a relativized substitution function and a variable assignment. In fact, the same substitution function \mathcal{I} (with parameters) can be used for all set-theoretic models; only the variable assignment needs to be varied. The substitution function \mathcal{I} is based on $\langle x, y \rangle \in v_1$ as $\sigma_{\in}(x, y)$, $\langle x, y \rangle \in v_2$ as $\sigma_{\text{Sat}}(x, y)$ and $x \in v_0$ as $\delta(x)$. No variable v'_n can be any of the three variables v_0, v_1 , and v_3 . The substitution function \mathcal{I} based on these three formulae looks as follows:

$$\mathcal{I}(\phi) := \begin{cases} \langle x', y' \rangle \in v_1 & \text{if } \phi \text{ is } x \in y, \\ x' = y' & \text{if } \phi \text{ is } x = y, \\ \langle x', y' \rangle \in v_2 & \text{if } \phi \text{ is } \text{Sat}(x, y), \\ \neg \mathcal{I}(\psi) & \text{if } \phi \text{ is } \neg \psi, \\ \mathcal{I}(\psi) \wedge \mathcal{I}(\chi) & \text{if } \phi \text{ is } \psi \wedge \chi, \text{ and} \\ \forall x' (x' \in v_0 \rightarrow \mathcal{I}(\psi)) & \text{if } \phi \text{ is } \forall x \psi \end{cases}$$

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

Given a variable assignment α over D and a model $\mathcal{M} := (D, E, S)$, the variable assignment $\alpha_{\mathcal{M}}$ is defined in set theory in the following way:

$$\begin{aligned} \alpha_{\mathcal{M}}(v_0) &:= D \\ \alpha_{\mathcal{M}}(v_1) &:= E \\ \alpha_{\mathcal{M}}(v_2) &:= S \\ \alpha_{\mathcal{M}}(v'_n) &:= \alpha(v_n) \end{aligned}$$

I state now the theorem showing that every set-theoretic model has a substitutional counterpart $(\mathcal{I}, \alpha_{\mathcal{M}})$.

Theorem 4.1 *For every set-theoretic model $\mathcal{M} := (D, E, S)$ there is a relativized substitution function \mathcal{I} that emulates all set-theoretic models in the following sense:*

$$\text{S} \vdash \forall \phi \forall \alpha \lambda D ((D, E, S) \models \phi[\alpha] \leftrightarrow \text{Sat}(\mathcal{I}(\phi), \alpha_{\mathcal{M}}))$$

As before, the quantifier $\forall \phi$ ranges here over all formulae of \mathcal{L}_{Sat} . Lemma 2.1 and the fact that $\mathcal{I}(\phi)$ is always in \mathcal{L} are used in the proof by induction on the length of ϕ .

Theorem 4.1 shows 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 $\alpha_{\mathcal{M}}$ plays a

dual role: On the one hand it supplies the values for the free variables v_0, v_1 , and v_2 in δ, σ_{\in} , and σ_{Sat} , on the other it contains the original variable assignment α .

Theorem 4.1 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 $(\mathcal{I}, \alpha_{\mathcal{M}})$ refutes the substitutional validity $\Gamma \models_S \phi$, where all elements of Γ and ϕ are \mathcal{L}_{Sat} -sentences:

Lemma 4.2 $\Gamma \models_S \phi$ implies $\Gamma \models \phi$

Here \models denotes the model-theoretic consequence relation with identity as logical constant.

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 [1967a] 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 and it can be used to show the equivalence of substitutional and model-theoretic validity. First, it is shown that some chosen calculus, say some Hilbert-style calculus, is sound with respect to substitutional validity:

Lemma 4.3 (soundness) $\Gamma \vdash \phi$ implies $\Gamma \models_S \phi$.

The upper index in \vdash indicates that the logical calculus contains axioms and/or rules for identity. The proof is by induction on the length of proofs and makes use of the axioms for $\text{Sat}(x, y)$ and relies on induction on a formula with the satisfaction predicate.

To complete the squeezing argument, the usual completeness theorem is invoked. That is, $\Gamma \models \phi$ implies $\Gamma \vdash \phi$. Combining this with lemmata 4.3 and 4.2 establishes that $\Gamma \models \phi$, $\Gamma \vdash \phi$, and $\Gamma \models_S \phi$ are all equivalent. In particular, model-theoretic and substitutional validity are equivalent:

Theorem 4.4 (equivalence of model-theoretic and substitutional consequence)

$$\Gamma \models_S \phi \text{ iff } \Gamma \models \phi$$

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

The right-to-left direction is obtained by observing that $\Gamma \models \phi$ implies $\Gamma \vdash \phi$ by the usual model-theoretic Gödel completeness theorem for predicate logic and then using the soundness lemma 4.3. The detour via provability and Gödel completeness can be avoided. Let a substitutional model (I, α) with a relativizing formula $\delta(x)$ be given. If the domain formula $\delta(x)$ of the substitutional model (I, α) defines a set, a corresponding set-theoretic model can easily be defined. If $\delta(x)$ is absent or does not define a set, one can show a lemma similar to the Löwenheim–Skolem downwards theorem which yields a substitutional set-sized ‘elementary submodel’ of (I, α) . 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 4.4 depends on the use of the untyped

axioms for Sat. Theorem 4.1 and therefore also lemma 4.2 could have been proved in a typed theory of satisfaction, that is, I 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. Type-free axioms could also be avoided in the soundness lemma 4.3 by admitting only substitution functions with a range containing only \mathcal{L} -formulae.

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 \mathcal{L} and not be universal. I retain universality by using a type-free theory of satisfaction.

4.4 Set-Theoretic Counterparts of Substitutional Models I have repeatedly claimed that, while for every set-theoretic model \mathcal{M} there is a corresponding substitutional model $\langle \mathcal{I}, \mathfrak{a}, \mathcal{M} \rangle$, the converse fails. This does not mean that there is a substitutional model $\langle I, \mathfrak{a} \rangle$ for which there is no set-theoretic model making the same sentences true, because that claim is obviously wrong: The Henkin completeness proofs for substitutional and model-theoretic validity tell us that for every consistent set of sentences there is a parameter-free substitutional as well as a model-theoretic model. Therefore, for every substitutional model $\langle I, \mathfrak{a} \rangle$, there is a set-theoretic model that makes the same sentences true:

$$\mathsf{S} \vdash \forall I \forall \mathfrak{a} \left(\text{SubFr}(I) \rightarrow \exists \mathcal{M} \forall \phi \left(\mathcal{M} \models \phi \leftrightarrow \text{Sat}(I(\phi), \mathfrak{a}) \right) \right)$$

Conversely, for every model-theoretic model there is a substitutional model that makes the same sentences true:

$$\mathsf{S} \vdash \forall \mathcal{M} \exists I \exists \mathfrak{a} \left(\text{SubFr}(I) \wedge \forall \phi \left(\mathcal{M} \models \phi \leftrightarrow \text{Sat}(I(\phi), \mathfrak{a}) \right) \right)$$

Hence, when I said that not every set-theoretic model corresponds to a substitutional model, I meant a stronger kind of correspondence. At least a corresponding set-theoretic model should interpret quantifiers in the same way as the substitutional model. They should be interpreted as ranging over the same objects in the following sense: Consider a substitutional model $\langle I, \mathfrak{a} \rangle$ with a relativized substitution function I based on the three formulae $\delta(x)$, $\sigma_{\in}(x, y)$, and $\sigma_{\text{Sat}}(x, y)$. Then the domain D of a corresponding set-theoretic model $\langle D, E, S \rangle$ should be $D = \{x : \delta(x)\}$. If, however, there is no set $\{x : \delta(x)\}$ or I does not feature a formulae $\delta(x)$, then there is no corresponding set-theoretic model. If the relativizing formula defines a set $D = \{x : \delta(x)\}$, then the corresponding set-theoretic model $\langle D, E, S \rangle$ is defined in S by taking the following extensions for \in and Sat:

$$\begin{aligned} E &:= \{ \langle x, y \rangle : \sigma_{\in}(x, y) \text{ and } x, y \in D \} \\ S &:= \{ \langle x, y \rangle : \sigma_{\text{Sat}}(x, y) \text{ and } x, y \in D \} \end{aligned}$$

Clearly, the homophonic substitutional or ‘intended’ interpretation has no set-theoretic counterpart; but a substitutional interpretation could also relativize quantifiers by restricting them to the constructible sets or ordinals. For such substitutional interpretations there is no corresponding set-theoretic model.

5 Extensions

5.1 Varying the Underlying Theories So far I have used the language \mathcal{L}_{Sat} as the language *for* which and the language *in* which logical validity is defined. My primary goal is to define logical validity for the language I am using; this language is the language of set theory. One can deviate from this setting and vary the object language for which validity is being defined, or the metatheory in which validity is defined, or both.

It may help to compare this with the model-theoretic account of logical validity: The usual model-theoretic definition of validity is carried out in set theory for *all* first-order languages (and other languages). The definition can also be given in much weaker theories and the completeness theorem be proved, depending on what exactly is required, for instance, whether uncountable languages are taken into account. Relevant results are Hilbert's and Bernays' [1939] formalized completeness theorem and the proof of the completeness theorem for suitable language in WKL_0 (an extension of an arithmetical system with Weak König's lemma) (see Simpson 24).

I have chosen set theory as the basis for \mathcal{S} in order compare substitutional validity with its main rival, the model-theoretic conception of logical validity. In particular, the choice of set theory makes it possible to show that for any model in the set-theoretic sense there is a corresponding substitutional model.

So far I have only defined validity for the language \mathcal{L}_{Sat} with only $=$, \in , and Sat as predicate symbols. One may wonder whether the definition of substitutional validity is as general as the model-theoretic and whether it can be given for arbitrary first-order languages, not only for \mathcal{L}_{Sat} . In fact it is not hard to define logical consequence for arbitrary first-order languages in \mathcal{S} . First we consider languages with other predicate symbols. Only the definition of a substitution function needs to be generalized: Every predicate symbol R with arity N is replaced with a formula $\sigma_R(x_1, \dots, x_n)$ of \mathcal{L}_{Sat} . Certain properties of substitution functions may be affected if languages with infinitely many nonlogical symbols are considered. With a finite vocabulary as in \mathcal{L}_{Sat} , substitution functions are always primitive recursive, because they need only specify finitely many formulae ($\sigma_{\in}(x, y)$, $\sigma_{\text{Sat}}(x, y)$ etc.). Obviously this is no longer the case with languages with an infinite vocabulary.

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 while avoiding variable clashes. Then the definition of logical consequence is modified to ensure that the constant always denotes 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:

$$\forall \Gamma \forall \phi \left(\Gamma \models_{\mathcal{S}}^{\equiv} \phi \ :\leftrightarrow \right. \\ \left. \forall I \forall \mathfrak{a} \left(\text{SubF}(I) \rightarrow (\text{Sat}(I(\Gamma), \mathfrak{a})) \wedge \text{Sat}(I(\ulcorner \exists x x = c \urcorner), \mathfrak{a}) \rightarrow \text{Sat}(I(\phi), \mathfrak{a})) \right) \right)$$

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

Function symbols with higher arity are less easy to deal with. It is not sufficient to substitute function symbols of arity n with functional expressions of arity n . We would like to be able to replace a given function symbol with an arbitrary function definable in our language, irrespective of whether it can be expressed using functions of our language or not. The problems are very similar to those encountered when defining the notion of relative interpretation for languages with function symbols. The methods used to define relative interpretations in these cases can be employed for the definition of substitutional validity for such languages as well.

All applications sketched so far do not require any changes to our axioms for the satisfaction predicate. Now I turn to the language *in* which logical consequence is defined. The change of language requires other axioms for satisfaction adapted to the new language, of course. It is possible to formulate the substitutional definition in other theories; this is often more straightforward than for the model-theoretic definition. For instance, instead of Zermelo-Fraenkel set theory, some sufficiently strong arithmetical theory or the weakened variant of ZF 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 assignments can be defined as functions with a subset of the set of variables as their domain. One can think of these functions still as assigning values to every variable by setting $\alpha(x)$ equal to some arbitrary object, say 0, if the variable x is not in the domain of the finite function. This facilitates the formulation of the Sat axioms. Otherwise, if $\alpha(x)$ is undefined for 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. The completeness theorem for substitutional consequence can be proved in weak theories. Once logical consequence is defined in weak theories, substitutional and model-theoretic (if these arithmetized versions still deserve this label) validity will look very similar, because the set-theoretic machinery of model theory has to be replaced with ‘more syntactic’ constructions.

5.2 Free Logic The substitutional definition of logical consequence with relativized substitution functions can be simplified by omitting the requirement that a relativizing formula has to be satisfied by some object. That is, substitution functions are defined as usual, but the definition of logical consequence is simplified as follows:

$$\forall \Gamma \forall \phi \left(\Gamma \models_S^{\text{free}} \phi \text{ :} \leftrightarrow \forall I \forall \alpha \left(\text{SubF}(I) \rightarrow (\text{Sat}(I(\Gamma), \alpha) \rightarrow \text{Sat}(I(\phi), \alpha)) \right) \right)$$

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. Consequently, the existence of at least one object becomes a logical truth. 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 [23] observed. Under the usual definitions there will not be any logical contradictions. Of course, there are very well-known workarounds and there are sensible definitions of truth in

a model, even when the empty domain is permitted. But I assume that the difficulties with the empty domain has led logicians to exclude the empty domain from the standard model-theoretic definition of logical validity. I consider the problems with the empty domain merely as a technical quirk of model theory. The empty domain should be admitted in the definition of logical validity; and it should not require any technical tricks. All problems with the empty domain disappear if the substitutional definition is employed. Once logical validity is defined substitutionally, free logic arises naturally.

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 $\text{Sat}(I(\ulcorner \exists x x = c \urcorner), \mathfrak{a})$. 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.

5.3 Other Logical Constants I have concentrated on classical first-order predicate logic. The reader might ask about other choices. If some other ‘logic’ is chosen such as intuitionistic, paracomplete or paraconsistent logic, the theory of satisfaction would have to be adapted. There are many possible choices. One could use a Kripke–Feferman like system, for instance, which provides a theory of nonclassical truth in classical logic. Such a theory of truth or satisfaction would naturally lead to a definition of validity in a nonclassical logic within a classical metatheory. Ideologically, I find such an approach not promising, because it clearly fails my desideratum of universality of logic. If one were serious about the nonclassical logic, then also the theory of truth and satisfaction should be used in the metatheory and a theory such as Partial Kripke–Feferman [10] could be employed. There are various difficulties, but in the end I expect that the substitutional approach can be applied to such a thoroughly non-classical setting.

There is a general point to be made: Truth theorists have compared the internal and external logics of truth theories. Roughly speaking, the internal logic are the sentences that are true according to the truth theory, while the external logic is the system in which the entire truth theory is formulated. If the notion of logical consequence is required to be universal, then the substitutional account requires a theory of truth and satisfaction in which internal and external logic do not significantly differ. Thus the substitutional account makes explicit why it is incoherent to reject an axiom such as (S3) $\forall \mathfrak{a} \forall \phi (\text{Sat}(\ulcorner \neg \phi \urcorner, \mathfrak{a}) \leftrightarrow \neg \text{Sat}(\ulcorner \phi \urcorner, \mathfrak{a}))$ and state the existence of truth value gaps or gluts, while *using* the rules for classical logic. I expect that certain philosophical ideas about truth theories, logical consequence, and semantic versus use theories of meaning can be made explicit with a substitutional analysis of validity based on an axiomatized satisfaction predicate.

Other readers may be happy with classical logic, but feel that logical constants have been omitted. In particular, it may be asked whether the substitutional account can be used for defining logical validity for languages with modal operators or generalized and higher-order quantifiers. Grossi [7] defined substitutional semantics for modal languages and proved adequacy results for standard model systems. Here I provide only some remarks about second-order logic. One could retain the theory S and define logic validity for second-order logic within S . For full second-order logic one would probably have to restrict the relativizing formulae $\delta(x)$ to those that define sets (for all interpretations of the parameters) and then assign only suitable sets to the second-order variables. Alternatively arbitrary $\delta(x)$ could be permitted and second-order variables replaced with formulae. Definitions of second-order validity along these lines are reductive in the sense that second-order logical consequence is defined in a first-order theory. If one is serious about second-order logic, S should be expanded to a second-order theory and the axiomatization of Sat would be extended to cover the second-order part of the expanded language. Of course, the first stumbling block is the non-axiomatizability of full second-order logic, but one could work in some second-order set-theoretic system with a primitive predicate for satisfaction. In such a framework, I expect that the definition of logical consequence can be given along the lines above. Only variable assignments would have to cover first- and second-order variables; otherwise no significant modifications would be required. Such an approach might be interesting for philosophers of set theory, because there would still be an intended interpretation for second-order set theory without any large cardinals or third-order quantifiers.

Notes

1. I admit that the relation of the present approach to early proofs and proof attempts of completeness theorems (such as Behmann's [2] for propositional logic) are not clear to me. In the absence of a model-theoretic notion of truth, they have a substitutional flavour. Of course they did not make use of a primitive truth predicate as I do; but perhaps they can be reconstructed using such a truth predicate.
2. The observation that Δ_2^0 -sets suffice for completeness appears in Kleene [14, p. 394, Theorem 35]. The history of this result dates further back. I am indebted to Walter Dean for hints to the literature. See his [4, p. 11].
3. Kreisel developed his *squeezing argument* in [15, 16, 17]. As a reason why model-theoretic validity is not obviously equivalent to intuitive validity, Kreisel mentions the absence of an intended interpretation for set theory. I am grateful to Göran Sundholm for making me aware of [15, p. 116f.] and [17, p. 253ff.]. For a more detailed exposition of the relation between the squeezing argument and my substitutional analysis see Halbach [9].
4. In Halbach [9] the term 'substitution function' was used for relativized substitution functions. A relativized substitution function replaces quantifiers $\forall v$ with quantifiers $\forall v(\delta(v) \rightarrow \dots)$ relativized to some formula $\delta(v)$. Relativized substitution functions are discussed below.
5. In an earlier version of the paper, I proved completeness indirectly, that is, via the completeness theorem for model-theoretic consequence. It relied on an observation similar to

theorem 4.1 below. A referee asked for a direct proof. The referee also wondered about the role of the parameters in the earlier completeness proof. The result above shows that completeness can be proved with a parameter-free substitution function, and thus parameters are not needed for proving the adequacy of \vdash .

6. Beall & Restall [1] call this definitional schema *Generalized Tarski Thesis*, although they prefer the more general and less specific sounding term *case* over *interpretation*. They credit Jeffrey [13] with the formulation. However, Tarski was just one of many authors who advocated a semantic definition, and he was not the first by far.
7. This phenomenon is well known from Etchemendy's [5] attack on Tarski's [26] definition of logical consequence; a related problem was spotted earlier by Hinman, Kim, & Stich [12] for a fixed finite domain. Williamson [31] and others have defended the view that these sentences are logical truths. McKeon [19] discusses the problem for Quine's account of substitutional validity. See also Wagner [30].
8. I am indebted to Anton Broberg for spotting the omission of such an axiom in earlier versions of the theory for satisfaction or truth. See [6] for a further discussion of this 'regularity' axiom.

References

- [1] Beall, J. and Restall, G. (2006). *Logical Pluralism*. Clarendon Press. 21
- [2] Behmann, H. (1922). Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen*, 86:163–229. 20
- [3] Berka, K. and Kreiser, L. (1986). *Logik-Texte: Kommentierte Auswahl zur Geschichte der modernen Logik*. Akademie-Verlag, Berlin, fourth, revised edition edition. First edition 1971, third, enlarged edition 1983. 22
- [4] Dean, W. (2019). Incompleteness via paradox and completeness. *Review of Symbolic Logic*. 20
- [5] Etchemendy, J. W. (1999). *The Concept of Logical Consequence*. CSLI, Stanford, second edition edition. First edition Cambridge (Mass.): Harvard University Press, 1990. 21
- [6] Fujimoto, K. and Halbach, V. (2019). Classical determinate truth. draft. 5, 21
- [7] Grossi, M. (2019). A substitutional semantics for modal logic. draft. 20
- [8] Halbach, V. (2014). *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge, revised edition. first edition 2011. 5
- [9] Halbach, V. (2018). The substitutional analysis of logical consequence. *Noûs*. 2, 3, 6, 20
- [10] Halbach, V. and Horsten, L. (2006). Axiomatizing Kripke's Theory of Truth. *Journal of Symbolic Logic*, 71:677–712. 19
- [11] Hilbert, D. and Bernays, P. (1939). *Grundlagen der Mathematik*, volume II. Springer, Berlin. 2, 17
- [12] Hinman, P. G., Kim, J., and Stich, S. P. (1968). Logical truth revisited. *Journal of*

- Philosophy*, 65:495–500. [21](#)
- [13] Jeffrey, R. C. (1992). *Formal Logic: Its Scope and Limits*. McGraw–Hill, third edition. [21](#)
- [14] Kleene, S. C. (1952). *Introduction to Metamathematics*. North Holland. [20](#)
- [15] Kreisel, G. (1965). Mathematical logic. In Saaty, T. L., editor, *Lectures on Modern Mathematics III*, pages 95–195. Wiley, New York. [20](#)
- [16] Kreisel, G. (1967a). Informal rigour and completeness proofs. In Lakatos, I., editor, *The Philosophy of Mathematics*, pages 138–171. North Holland, Amsterdam. [15](#), [20](#)
- [17] Kreisel, G. (1967b). Mathematical logic: what has it done for the philosophy of mathematics? In Schoenman, R., editor, *Bertrand Russell: Philosopher of the Century*, pages 201–272. George Allen & Unwin, London. [20](#)
- [18] McGee, V. (1992). Maximal consistent sets of instances of Tarski’s schema (T). *Journal of Philosophical Logic*, 21:235–241. [2](#)
- [19] McKeon, M. (2004). On the substitutional characterization of first-order logical truth. *History and Philosophy of Logic*, 25:205–224. [21](#)
- [20] Quine, W. V. (1986). *Philosophy of Logic*. Harvard University Press, second edition. [1](#)
- [21] Rayo, A. (2006). Beyond plurals. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 220–254. Oxford University Press. [2](#)
- [22] Rayo, A. and Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In *Liars and Heaps: New Essays on Paradox*, pages 331–356. Oxford University Press. [12](#)
- [23] Schneider, H. H. (1958). Semantics of the predicate calculus with identity and the validity in the empty individual-domain. *Portugaliae Mathematica*, 17:85–96. [18](#)
- [24] Simpson, S. (1998). *Subsystems of Second Order Arithmetic*. Springer, Berlin. [17](#)
- [25] Studd, J. P. (2019). *Everything, More or Less: A Defence of Generality Relativism*. Oxford University Press. [2](#)
- [26] Tarski, A. (1936). Über den Begriff der logischen Folgerung. *Actes du congrès international de philosophie scientifique*, 7:1–11. Reprinted (abbreviated) in [\[3\]](#), 404–13]; also reprinted as “On the Concept of Logical Consequence” in [\[27\]](#), 409–20]; page references are given for the translation. [3](#), [11](#), [12](#), [16](#), [21](#)
- [27] Tarski, A. (1983). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Hackett Publishing, Indianapolis, second edition edition. [22](#)
- [28] Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, 7:143–154. edited by John Corcoran. [3](#)
- [29] Tarski, A. and Vaught, R. (1956). Arithmetical extensions of relational systems. *Compositio Mathematica*, 13:81–102. [3](#)
- [30] Wagner, H. (2019). Quine’s substitutional definition of logical truth and the philosophical significance of the Löwenheim–Hilbert–Bernays theorem. *History and Philosophy of Logic*, 40:182–199. [21](#)

- [31] Williamson, T. (1999). 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). [2](#), [3](#), [21](#)

Acknowledgments

I am grateful to members of various audiences of talks on topics presented in this paper. In particular, I am indebted to Anton Broberg, Walter Dean, Kentaro Fujimoto, Leon Horsten, Øystein Linnebo, Beau Mount, Carlo Nicolai, Lavinia Picollo, Göran Sundholm, and Timothy Williamson. I thank a referee for thoughtful comments, which prompted me to rethink central issues and rewrite the paper.

University of Oxford
New College
OX1 3BN Oxford
England
volker.halbach@new.ox.ac.uk
<http://users.ox.ac.uk/~sfop0114/>