THE SUBSTITUTIONAL ANALYSIS OF LOGICAL CONSEQUENCE

Volker Halbach*

9th July 2016

> Consequentia 'formalis' vocatur quae in omnibus terminis valet retenta forma consimili. Vel si vis expresse loqui de vi sermonis, consequentia formalis est cui omnis propositio similis in forma quae formaretur esset bona consequentia [...]
>
> Iohannes Buridanus, Tractatus de Consequentiis (Hubien 1976, 1.3, p.22f)

ABSTRACT

A substitutional account of logical truth and consequence is developed and defended against competing accounts such as the model-theoretic definition of validity. Roughly, a substitution instance of a sentence is defined as the result of uniformly substituting nonlogical expressions in the sentence with expressions of the same grammatical category and possibly relativizing quantifiers. In particular, predicate symbols can be replaced with formulae possibly containing additional free variables. A sentence is defined to be logically true iff all its substitution instances are satisfied by all variable assignments. Logical consequence is defined analogously. Satisfaction is introduced axiomatically. The notion of logical validity defined in this way is universal, that is, validity is not defined in some metalanguage, but rather in the language in which validity is defined. It is shown that for every model-theoretic interpretation there is a corresponding substitutional interpretation in a sense ot be specified. Conversely, however, there are substitutional interpretations – in particular the trivial intended interpretation – that lack a model-theoretic counterpart. Thus the definition of substitutional validity overcomes the weaknesses of more restrictive accounts of substitutional validity. In Kreisel's squeezing

*I am grateful to Walter Dean, Beau Mount, Carlo Nicolai, Thomas Schindler, Göran Sundholm, Albert Visser, Philip Welch and Timothy Williamson for discussions about the themes of this paper. I also thank the audiences of talks about this material at FilMat2016 at Chieti, the PhilMath Intersem meeting on 7th June 2016, and the Bristol-München Conference on Truth and Rationality 2016.

argument the formal notion of substitutional validity naturally slots into the place of informal intuitive validity.

Formal validity

At the origin of logic is the observation that arguments sharing certain forms never have true premisses and a false conclusion. Similarly, all sentences of certain forms are always true. Arguments and sentences of this kind are *formally* valid. From the outset logicians have been concerned with the study and systematization of these arguments, sentences and their forms. For instance, arguments in *modus barbara* are formally valid:

> All men are mortal. Socrates is a man. Therefore Socrates is mortal.

If the terms 'man', 'Socrates' and 'mortal' are uniformly replaced with other terms of the same grammatical category, the resulting argument will never have true premisses and a false conclusion.

There are arguments in which the truth of the premisses necessarily implies the truth of the conclusion, but which are not *formally* valid, at least not by he usual standards. In particular, an argument may be *analytically* valid without being formally valid. An example is the following argument:

> John is a bachelor. Therefore John is unmarried.

Although the following argument is not analytically valid, its conclusion is necessarily implied by the premiss under common assumptions expounded by Kripke (1972):

> There is $\mathsf{H_2O}$ in the beaker. Therefore there is water in the beaker.

Arguments of this kind could be called *metaphysically* valid.

Since analytically and metaphysically valid arguments are not formally valid, necessary truth preservation is not a sufficient condition for formal validity. It could be objected that this depends on a certain understanding of necessity: If we understand necessity as formal or logical necessity, then necessary truth preservation is a sufficient condition for formal validity. However, tweaking the notion of necessity in this way runs the risk of becoming circular: Logical necessity cannot be analyzed again as formal validity. Hence it seems hard to avoid the conclusion that there are necessarily truth preserving arguments that are not formally valid.

Many medieval philosopher at least from Buridan onwards were clear about the distinction between formal validity and other kinds of validity, although not all used the term 'formal' in the same way.[1] More recently, philosophers have usually relegated the analysis of analytic, metaphysical and other kinds of 'material' validity to philosophy of language. Logical validity has come to be understood as formal validity.[2] In what follows I identify logical with formal

[1] See (Read 2012) and (Aho and Yrjönsuuri 2009, §6.3).

[2] See Asmus and Restall (2012) for a brief historical synopsis.

validity, too. The concentration on formal validity has allowed logicians to purge logic from non-formal modal notions. On the model-theoretic and the substitutional account, logical validity and necessity are disentangled.[3]

THE SUBSTITUTION CRITERION

Counterexamples have always played an important role in arguing that a given argument fails to be logically valid. Traditionally, counterexamples are conceived as substitution instances. This is still the way counterexamples are often presented in introductory logic classes. A substitution instance of a given argument is obtained by substituting uniformly nonlogical (or, in a more traditional terminology, categorematic) terms with nonlogical terms of the same grammatical category. A substitution instance of an argument is a counterexample if and only if the premisses of the counterexample are true and the the conclusion is false.

The possibility of using of counterexamples for showing that an argument is not logically valid relies on the following soundness principle:

> SOUNDNESS If an argument is logically valid, it does not have any counterexamples.

The converse of this principle is more problematic. It can be stated as follows, by contraposition, as the following completeness principle:

> COMPLETENESS If an argument is not logically valid, it does have a counterexample.

The existence of substitutional counterexamples depends on the availability of suitable substitution instances in the language. Thus the completeness principle seems to make logical validity highly dependent on the language from which the substitution instances can be taken. In particular, if certain objects and relations cannot be expressed in the language, sentences may be classified as valid although they are not.[4] This worry is closely related to the *persistence* problem: If new vocabulary is added, then, so is argued, sentences that have been previously classified as valid may become invalid. I will show that difficulties of this kind do not arise if

[3] There are arguments against the identification of formal and logical validity. See, for instance, (Read 1994) for a criticism of the identification. Modal arguments play a crucial role in Etchemendy's (1990) attack on the model-theoretic definition of logical consequence. They have been countered by McGee (1992) and others. The debate is also relevant for the substitutional account. However, here in this paper the connection between modality and logical consequence will not be discussed.

[4] This was the main reason for Tarski (1936*b*) to reject a substitutional analysis, which he discussed on p. 417. He stated a substitutional definition of logical consequence as 'condition (F)' and then went on: 'It may, and it does, happen – it is not difficult to show this by considering special formalized languages – that the sentence X does not follow in the ordinary sense from the sentences of the class K although the condition (F) is satisfied. This condition may in fact be satisfied only because the language with which we are dealing does not possess a sufficient stock of extra-logical constants. The condition (F) could be regarded as sufficient for the sentence X to follow from the class K only if the designations of all possible objects occurred in the language in question. This assumption, however, is fictitious and can never be realized.' I will argue that this objection can be countered with a suitable substitutional definition of logical consequence.

substitution instances of sentences are suitably defined and, in particular, allowed to contain free variables.

Combining both soundness and completeness yields the following criterion for logical validity:

> SUBSTITUTION CRITERION An argument is logically valid if and only if it has no counterexamples.

This principle has been used not only as a criterion but as a conceptual analysis of logical consequence.[5]

In the present paper I explore the potential of the substitutional account of logical truth and consequence for highly regimented languages, more precisely, for strong classical first-order theories such as set theory in which large parts of mathematics and, perhaps, the sciences can be developed. I have little to say on the subtleties of natural language. The reader more interested in logical consequence in natural language may take first-order languages as a test case. If the substitutional theory is successful for them, there is at least hope it can be extended to natural language.

TERMINOLOGICAL REMARK. I apply the terms 'logical validity' and 'formal validity' to sentences and arguments. When I say that a sentence is valid, this may be understood as the claim that the argument with the empty premiss set and the sentence as conclusion is valid. Occasionally the qualification 'logically' or 'formal' is omitted if there is no risk that logical validity is confused with validity in a model or other kinds of validity. For single sentences I use the term 'logical truth' as synonymous with 'logical validity'. The expression 'logical truth' suggests that it designates a special kind of truth and that logical truth is defined as truth with some extra condition. Although this understanding is in line with my own approach, it should not be assumed from the outset that truth is to be defined from an absolute notion of truth that is not relative to models and an additional condition that makes it 'logical'.

[5] The substitution principle seems to have been used early on, at least implicitly. The quote from Buridan's *Tractatus de Consequentiis* used as motto of this paper plus parts of the following paragraph on material consequence come at least close to an endorsement of the substitution criterion as a definition of formal consequence. See (Dutilh Novaes 2012) and (Aho and Yrjönsuuri 2009, §6.3) for more details. In what follows I refer to this analysis as the substitutional definition or conception of logical consequence

According to the usual narrative, the substitutional notion of validity finally developed into the modern model-theoretic account of logical consequence via Bolzano's (1837) and Tarski's (1936*b*). Kreisel (1967*a*) and Etchemendy (1990) both mentioned Bolzano as a precursor of Tarski's and the contemporary model-theoretic definition of logical validity. In particular, (Etchemendy 1990, p. 28ff) saw Bolzano as a proponent of a linguistic substitutional account of logical truth In footnote 2 to chapter 3 Etchemendy made some qualifications, but then says that he will 'gloss over this difference'. Ironically, Tarski (1936*b*) added later a footnote to his paper mentioning an observation by Heinrich Scholz with Bolzano's account as precursor of the definition of logical consequence advocated in Tarski's paper and not the substitutional one. Bolzano is not the best example of a proponent of a substitutional account. In Bolzano what is substituted are not linguistic entities but rather what Bolzano called 'Vorstellung'. This German term is usually translated as 'idea'. These *Vorstellungen* are neither linguistic nor psychological entities. So Tarski's view of Bolzano's theory may be more accurate than Etchemendy's. Medieval logicians like Buridan provide much better examples of substitutional theories. There are also have been attempts to revive the substitutional approach, in particular by Quine (1986).

For formalized languages the substitutional account of logical validity has largely been superseded by the proof-theoretic and the model-theoretic approaches. According to the proof-theoretic or inferentialist conception, roughly, an argument is valid if and only if the conclusion can be derived from the premisses using certain rules and axioms, very often, the rules of Gentzen's system of Natural Deduction. Similarly, a sentence is valid if and only if the sentence is derivable without any premisses.

The model-theoretic analysis is closer to the substitutional. They have a common form; both can be stated in the following way:

> *Generalized Tarski Thesis* An argument is logically valid iff the conclusion is true under all interpretations under which its premisses are true.[6]

I call analyses of this form *semantic* to distinguish them from proof-theoretic accounts that rely neither on interpretations nor a notion of truth. The main difference between a substitutional and the model-theoretic analysis lies in the notion of interpretation: On a substitutional approach, an interpretation is understood in a syntactic way as a function replacing nonlogical terms; on the model-theoretic approach an interpretation is an assignment of semantic values to the nonlogical expressions plus the specification of a domain. There is another important difference: The model-theoretic analysis of validity relies on a set-theoretic definition of truth in a model. The substitutional account, in contrast, requires an 'absolute' notion of truth that is not relativized to a (set-sized) model, as will be shown below.

The substitutional account was developed into the model-theoretic definition of logical consequence mainly by Tarski, starting with his (1936*b*). The modern proper notion of truth in a model appeared only in (Tarski and Vaught 1956).[7]

Although a lot of effort has been spent to defend inferentialist and model-theoretic accounts of logical consequence, there is a widespread suspicion that neither is adequate. There is a fundamental reason for this suspicion: The proof-theoretic and the model-theoretic definitions do not directly capture the central feature of logical consequence, that is, truth preservation. That is, it is not directly obvious that a valid argument cannot have true premisses and a false conclusion. The inferentialist analysis is not semantic and does not mention truth at all. In semantic theories, that is, theories conforming with the Generalized Tarski Thesis, truth preservation seems to be a built into the definition of logical consequence. For the model-theoretic account of logical consequence, however, it is a well-known embarrassment that truth preservation is not an obvious property of logically valid arguments and that truth is not an obvious property of logical truths. Truth preservation does not trivially follow

[6] Beall and Restall (2006) prefer the more general and less specific sounding term *case* over *interpretation*, when it comes to give a general form of the definition their Generalized Tarski Thesis. They credit Jeffrey (1992) with the formulation.

[7] Only with the modern notion of a model with a domain a universal definition of validity can be given. In Tarski's (1936*b*), in contrast, logical consequence is defined only for an objectlanguage in a stronger metalanguage. With the modern notion logical validity for the set-theoretic language can be defined within set theory.

from the model-theoretic definition of logical consequence, because the universal quantifier over interpretations in the definition ranges over all models in the technical sense; but the 'intended interpretation' is not one of these models and cannot easily be identified with one of these models. Models have set-sized domains, while the intended interpretation, if it could be conceived as a model, cannot be limited by any cardinality. Similarly, logical truth defined as truth in all models does not imply truth simpliciter.

If logical truth is understood as truth under all interpretations, then it is an oddity of the model-theoretic account that there is not a special interpretation of a given sentence that allows one to understand the sentence at face value. Several philosophers and logicians have felt the need for such an intended interpretation conceived as an intended model. In a set-theoretic setting, which is usually assumed and also below in the present paper, the intended model would be the set-theoretic universe V. Of course there are also other motives for talking about the 'set-theoretic universe'. Most set-theorists will happily talk about the set-theoretic universe, a cumulative hierarchy, and so on. There are ways to make this precise in a class theory and declare the universe to be a proper class about which we can reason in our preferred class theory. However, as is well known, this move will not help. We can start with a class theory as base theory. But then the intended interpretation would have to contain all sets and, in addition, proper classes. Such objects can be talked about in a theory about superclasses and a type theory over set theory can be set up. This does not look very appealing; and, more importantly, one will never reach the elusive intended interpretation.

A kind of dual approach is to resort to the indefinite extensibility of the set-theoretic universe. The approach is hard to describe consistently. Very roughly speaking, it is denied that quantifiers of the object language range over all sets. The universal quantifier ranges over a set-sized universe, a set V_α, that is, a level in the cumulative set-theoretic universe; but this set is then contained in another level and so on. Of course this is very similar to postulating classes, superclasses and further levels.

Neither class theory, the impossibility of absolutely general quantification, nor indefinite extensibility seem to be attractive ways of addressing the problem of the 'missing' intended model. It is not only that set theory does not prove the existence of the set-theoretic universe. Standard Zermelo–Fraenkel set theory *refutes* the existence of a universal set. Here I do not dwell on the philosophy of set theory. But it strikes me as very strange that the theory of logical consequence should push us into the most extravagant metaphysical speculations about universes and intended interpretations that slip out of sight like Berkeley's elusive subject as soon as we think we have snatched a glimpsed of it.

On the substitutional account of logical validity, the need for all this controversial and suspect theorizing disappears. The intended interpretation is no longer mysterious or elusive. It becomes a trivial function that just maps every sentence to itself. Under the intended interpretation a sentence is understood at face value; it is not reinterpreted in any way. There is no longer a need for trying to form a universal object or 'reality'. Since the intended interpretation in the sense of the substitutional account can be used in a completely unproblematic way, the proof that logical truth implies truth becomes trivial, and as does the proof that logical consequence is truth preserving. This is just what one should expect from an adequate conceptual analysis of validity.

Because truth in all set-theoretic models does not imply the truth in the intended model, some logicians have tried to come up with arguments for the claim that at least the extension of the notion of model-theoretic validity coincides with that of 'intuitive' validity. In particular Kreisel (1965, 1967*a*,*b*)[8] employed his squeezing argument to show that the two formal analyses coincide with the somewhat elusive 'intuitive' notion of validity. I sketch the argument for logical truth; it applies also to logical consequence in a straightforward manner.

According to the squeezing argument, provability of a sentence in a suitable deductive system ('$\vdash_{\mathrm{ND}} \phi$' in the diagram below) implies its intuitive validity. The intuitive validity of a sentence in turn implies its model-theoretic validity, because any model-theoretic counterexample also refutes the intuitive logical truth of a sentence. These two implications – both straight arrows in the diagram – cannot be proved formally, because the notion of intuitive validity is not formally defined. However, there is a formal theorem, the completeness theorem for first-order logic, that shows that the model-theoretic validity of a sentence implies its provability in the logical calculus. The implications are shown in the following diagram:

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

If the two informal implications visualized by the two straight arrows in the diagram hold, all three notions have the same extension. For the mathematical logicians fixing the extension of logical validity is usually good enough. But the squeezing argument does not establish that any of the two formal definitions is an adequate conceptual analysis of logical validity. This situation strikes me as highly unsatisfactory: We have at least two formally precise characterizations of the extension of the notion of logical validity without having an adequate conceptual analysis. Logical consequence is not a pretheoretical concept. The notion of logical consequence especially for formal languages has been honed with great rigour. It would be very strange if such a theoretical notion eluded any attempt to make it formally precise.

I put forward the substitutional analysis as a direct, explicit, formal, and rigorous analysis of logical consequence. The substitutional definition of logical validity, if correctly spelled out, slots directly into the place of 'intuitive validity' in Kreisel's squeezing argument. The substitutional account does not suffer from the main problems of the proof- and model-theoretic accounts.

First, substitutional validity is closer to rough and less rigorous definitions of validity as they are given in introductory logic courses. It is also closer to how logicians over the centuries have specified counterexamples and established validity before finally the model-theoretic account became prevalent.

[8] I am indebted to Göran Sundholm for making me aware of (Kreisel 1965, p. 116f.) and (Kreisel 1967*b*, p. 253ff.).

Secondly, on a substitutional account it is obvious why logical truth implies truth *simpliciter* and why logical consequence is truth preserving. The 'intended interpretation' is completely trivial on the substitutional account: It is the identity function on formulae that does not replace anything.

Thirdly, the substitutional definition of logical consequence is not tied to set theory and its philosophy. On the model-theoretic account, interpretations are specific sets; on the substitutional account they are merely syntactic and (under certain natural assumptions) computable functions replacing expressions and perhaps an assignment of objects to the free variables. As mentioned above, work on the model-theoretic theory of logical consequence has led philosophers to doubt that quantification 'over absolutely everything' is possible and to speculate about the indefinite extensibility of the set-theoretic universe. For the proponent of a substitutional account it is easier to avoid such speculations. At any rate the direct link between the most complex metaphysical speculative theory hitherto, set theory, and the theory of logical consequence is severed.

THE CHOICE OF PRIMITIVE NOTIONS

I will provide a definition of logical validity in a base theory – with set theory being the main example – expanded with a primitive, axiomatized predicate for satisfaction. This is contrast to the usual model-theoretic analysis that does not require a primitive predicate for satisfaction, because satisfaction in a model can be defined in set theory alone. For the substitutional analysis set-theoretic reductionism has to be abandoned: The substitutional notion of validity is not reducible to set theory alone.[9]

If no primitive notion beyond set theory itself are admitted, I cannot see how to avoid the problems of the model-theoretic analysis sketched above and several other problems identified by Etchemendy (1990). Basic properties of logical validity become mysterious on a strongly reductive approach that excludes all notions that are not purely set-theoretic. In particular, I cannot see how to obtain a definition of the logical validity of arguments that immediately entails truth preservation in valid arguments, as it is not even clear how to *state* properties such as truth preservation.

If set-theoretic reductionism has to be abandoned and primitive notions beyond set theory have to be used, then why not go all the way and treat validity as a primitive predicate?[10] First, satisfaction will be required anyway for various areas of philosophy. For instance, in epistemology, truth is needed to formulate the fairly uncontroversial claim that the truth of a belief is a necessary condition for it to be known. Logical validity is less entrenched in

[9] This applies at least to the definition of substitutional validity defended here. In footnote 19 I sketch a (less satisfactory) definition of substitutional validity that doesn't require an additional truth predicate and is compatible with set-theoretic reductionism.

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

various philosophical disciplines. Secondly, we cannot easily appeal to a intuitive pretheoretic notion of logical consequence. The notion of logical is theoretical and usually introduced by a characterization in the style of the Generalized Tarski Thesis above. Truth is thus conceptually prior to truth. Thirdly, on the approach envisaged here, logical validity is definable in terms of a satisfaction predicate and other weak resources. Conversely, however, satisfaction cannot be defined in terms of logical validity and set theory. If logical validity is treated as a primitive notion, truth or satisfaction still would have to be added as primitive notion, in order to show truth preservation and other desired properties of logical validity. That satisfaction and truth are not definable for first-order languages in terms of logical validity is to be expected for purely recursion-theoretic grounds: Logical validity will be extensionally equivalent to first-order provability and thus be recursively enumerable, while truth is not even be elementarily definable.

Hidden in the definition of a substitution instance is another crucial notion, the notion of a logical constant, as will become obvious below. All the analyses of logical validity mentioned so far rely on a distinction between logical and nonlogical vocabulary. In the present paper I would like to avoid making any strong commitments to a particular way of distinguishing between logical and nonlogical vocabulary. I apply a rather traditional way of making the distinction, as will become apparent below.

Most of the various criteria that have been used to make the distinction can be applied also on a substitutional account. If they are formulated in a model-theoretic framework, such as the permutation criterion, some adjustments will have to be made; but such criteria can often be rephrased in substitutional terms. I leave the discussion for another occasion.

SUBSTITUTION INSTANCES

Even if the logical terms are fixed, it may not be clear what counts as a suitable substitution instance. A suitable definition of substitution instances will be the key to the solution of the persistence and completeness problem. To motivate the definition for formalized languages, I look at examples in natural language.

Clearly, a general term such as 'is a man' can be replaced not only with another general term but also with a complex general term such as 'is a wise philosopher with a long beard'. The argument

> All wise philosophers with a long beard are mortal. Socrates is a wise philosopher with a long beard. Therefore Socrates is mortal.

is of the same form *modus barbara* as the usual Socrates–man example. Similarly, we may want to allow atomic singular terms to be replaced with complex ones. Replacing proper names with definite descriptions may cause problems; and this requires some care. I leave the elaboration of the formal details to another occasion.

There are more singular terms than only proper names and definite descriptions. Would we allow personal or demonstrative pronouns as substitution instances of proper names? For instance, is the following argument of the same form as the above argument in *modus barbara*?

> All starfish live in the sea. That animal is a starfish. Therefore that animal lives in the sea.

The phrase 'that animal' can also be replaced with the mere pronoun 'that', even though that may sound less idiomatic. At any rate, the resulting argument is of the same form in *modus barbara* and logically valid again, as long as all occurrences of 'that animal' or just 'that' refer to the same object. Whenever both premisses are true, the conclusion is true, whatever 'that' refers to. Thus pronouns can be substituted for singular terms.[11]

Pronouns can also be introduced through the substitution of predicate expressions.

> All objects in the box are smaller than that (object). The pen is in the box. Therefore it is smaller than that (object).

Here 'is smaller than that' has replaced 'is mortal'. The form of the argument has not changed, as long as the reference of 'that' does not change between the premisses and the conclusion. Since occurrences of pronouns in substitution instances are going to be allowed, the definition of logical truth in natural language would require a reference to the way the pronouns are interpreted:

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

The use of demonstrative or personal pronouns makes it possible to formulate counterexamples or interpretations involving a singular term referring to an object for which we lack a name or definite description.

SUBSTITUTIONAL INTERPRETATIONS

The observations from the last section will now be applied to a formal language. In a language of first-order logic free variables can play the role of pronouns. A formula with free variables is not true simpliciter; it is only true relative to the reference of the free variables. The reference of the free variables can be specified by a variable assignment. Consequently, the formally precise definition of validity will take the following form: A sentence is logically valid iff all substitution instances are satisfied by all variable assignments. Therefore not only a notion of truth but also of satisfaction will be required.

Substitution instances of a sentence or a set of sentences will be defined as the result of applying a substitutional interpretation to the sentence or the set. A substitutional interpretation, in turn, will be defined as a function that yields, applied to a formulae of the language (including those with the satisfaction predicate), a substitution instance.[12] The substitution

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

[12] The term 'substitutional interpretation' may be somewhat misleading as it usually refers to a particular interpretation of quantifiers. This may suggest that the nonlogical vocabulary is quantified away and logical truth defined in terms of higher order quantification. This approach is not completely unrelated, but here it is not required because in the presence of a satisfaction predicate higher-order quantifiers are not required.

will be uniform, of course. That is under a given substitutional interpretation, the same predicate symbol, for instance, will always be replaced with the same formula. Substitutional interpretations resemble relative interpretations as introduced by Tarski et al. (1953). Roughly, substitutional interpretations are defined like relative interpretation just without the requirement that provability is preserved.

A substitutional interpretation maps all formulae of the entire language to formulae in such a way that the logical structure is preserved. If there are no function symbols in the formula, then a substitutional interpretation is a function that replaces uniformly in every given formula each atomic formula with a possibly complex formula; moreover, it possibly replaces every quantifier with a quantifier restricted to some fixed formula (or it may leave the quantifiers unchanged). Clashes of variables are assumed to be avoided in some of the usual ways. Individual constants are uniformly replaced with other constants or variables. If further function symbols are present, things become more complicated; I do not go into details here.

Here I do not take a stance on the logicality of identity. We can say that a substitutional interpretation does not replace any occurrence of the identity symbol, so that identity is treated as a logical constant. There are some further tweaks of the definition of substitutional interpretations that will be considered in the following section.

The technical details of defining substitutional interpretations recursively are not difficult. Obviously the definitions can be carried in weak systems already. Assuming that there are only finitely many nonlogical symbols, substitutional interpretations are primitive recursive functions.

THE QUANTIFIERS

In the definition of substitutional interpretations there is an issue whose direct analogue on the model-theoretic side is the question whether the model with the empty domain should be admitted.

As explained above, a substitutional interpretation restricts all quantifiers in a formula. So, under a substitutional interpretation I, a subformula $\forall x\, F(x)$ becomes $\forall x\, (R(x) \rightarrow I(F(x)))$ and $\exists x\, F(x)$ becomes $\exists x\, (R(x) \wedge I(F(x)))$. The restricting formula $R(x)$ corresponds to the domain of a model-theoretic interpretation, that is, of a model.

I have not ruled out relativizing formulae such as $P(x) \wedge \neg P(x)$ that do not apply to anything. This corresponds to an empty domain on the model-theoretic approach. Of course, empty domains are not admitted in the standard semantics of classical first-order logic. If this is to be emulated on the substitutional approach, the definition of substitutional interpretations can be changed as follows: The substitutional interpretation $I(F)$ with relativizing formula $R(x)$ of a formula F containing, for example, exactly the individual constants a and b is preceded by the expression $\exists x R(x) \wedge R(a) \wedge R(b) \rightarrow \ldots$. The formula $\exists x R(x)$ expresses, model-theoretically speaking, that the domain is not empty, while $R(a)$ ensures that the constant a denotes an object in the domain.

The exclusion of the empty domain in models strikes me as a philosophically not very convincing oddity that is accepted mainly for convenience. One difficulty with free logic is the definition of truth from satisfaction. As Schneider (1958) noticed, if a sentence is defined to be true iff it is satisfied by all variable assignments, then all sentences will be true in the model with the empty domain, because there are no variable assignments over the empty domain. If a sentence is defined to be true iff it is satisfied by at least one variable assignment, no sentence is true. The problems can be solved. Williamson (1999) discussed some workarounds; but for most purposes it is much more convenient just to exclude the empty domain. Usually logicians consider theories that imply existential claims. Thus the empty domain is excluded by nonlogical axioms and the model with the empty domain is irrelevant for the analysis of these theories.

For the analysis of logical validity, however, there are good reason to retain the empty domain, if validity is analyzed in the model-theoretic way, even if this means that the definitions of satisfaction and truth become more cumbersome. On the substitutional account the analogous move is much more straightforward. On the contrary, free logic looks much more natural, because the antecedent including $\exists x R(x)$ of each interpretation can be omitted. However, there are still decisions to be made as far as individual constants are concerned. If negative free logic is chosen, then the conjunct $R(t)$ (with t a variable or individual constant) should be adjoined to the substitutional interpretation of an atomic formula $F(t)$, where $R(x)$ is the relativizing formula of the substitutional interpretation. This ensures that formulae with constants not denoting objects satisfying $R(x)$ are not satisfied by any variable assignment.

If a free logic approach is chosen, one may think about adapting the logic of the base theory and, in particular, the treatment of individual constants. The necessary modifications are straightforward. If the language of the base theory does not contain any individual constants, there is no need for any modifications, as the base theory will contain existential claims.

Here I do not discuss the pros and cons of the different varieties of free logic. Many arguments that have been made in favour and against certain varieties in terms of models can be rephrased in terms of substitutional interpretations. If no modifications are made to the definition of substitutional interpretations in the previous section, then sentences of the form $\exists x(P(x) \vee \neg P(x))$ will not come out as logically valid under the definitions of logical truth in the next section. The notion of validity will be that of free logic. The most straightforward definition of validity on the substitutional account yields a logic without any ontological commitment. I take this to be an advantage of the substitutional analysis. No additional trickery is required.

On the model-theoretic account it is not straightforward to accommodate the empty domain; but the opposite extreme is even more difficult to handle, that is, quantification over everything. What cannot be so easily changed on the model-theoretic account is the way quantification is treated. In a sense quantifiers are not interpreted in the same way in all models. The interpretation of the quantifiers is provided by the domain of a model; and this domain varies from model to model. A domain cannot contain all objects, because a domain has to be a set, not a proper class. Williamson (2000) and others have rejected the restriction to a domain of quantification. They treat quantifiers as logical symbols with constant meaning across all interpretations. This means that sentences expressing that there are at least n many objects become logically valid. But it is not so easy to dispense with the domain. If the clauses for the quantifiers in the definition of truth are not restricted to sets (as suggested by Williamson 2000, p. 326), then the inductive definition of satisfaction cannot be shown to have a fixed point; that is, the inductive definition cannot be converted into an explicit definition. Of course, the definition of satisfaction could be carried out in a metatheory that allows us to talk about proper classes as objects. But that would just mean that the quantification in the objectlanguage is restricted to sets while classes are excluded. Hence the restriction to domains is very much built into the model-theoretic definition of satisfaction and cannot easily be eliminated from it. The substitutional account, in contrast, is not confined in this way. Hence if quantifiers should be constantly be interpreted as ranging over everything, this can be easily achieved by dropping the restricting formulae in substitutional interpretations.

SATISFACTION AS A PRIMITIVE NOTION

A sentence is defined to be logically valid iff all its substitution instances are satisfied by all variable assignments. The trivial substitution that maps a sentence to itself is a permissible substitution. Thus, if we have the notion of truth or satisfaction relative to a substitution, we can define an 'absolute' notion of truth as truth, that is satisfaction under all variable assignments, relative to the trivial substitution. By Tarski's theorem on the undefinability of truth, such an absolute notion is not definable. Since the substitutional account of logical validity requires a notion an absolute notion of truth, which cannot be defined, I introduce a notion of satisfaction axiomatically.

It is not necessary to axiomatize a notion of satisfaction relative to a substitution. Given an 'absolute' notion of truth in the sense of (Davidson 1973), truth relative to a substitution can be defined, as substitution is a syntactic and computable concept that is definable in weak arithmetical systems already, as long as substitutional interpretations are defined in a straightforward way and the languages are not grotesquely complicated.

Axiomatic approaches to truth and satisfaction have been pursued and advocated by different authors.[13] In this paper the axioms for satisfaction will be added to an extension of Zermelo–Fraenkel set theory possibly with urelements as the 'base theory'. The base theory can be enriched by further defined notions and further axioms and rules may be added. One could also use weaker theories as base theory such as certain arithmetical theories, but then some adjustments will be required. The base theory must contain a theory of syntax. This can be achieved in the usual way by a coding or by a direct axiomatization of syntax. Furthermore, it must contain a theory of variable assignments, that is, functions from the set of variables into arbitrary sets. With a a little extra work one can use also finite functions as variable assignments. This will be necessary if an arithmetical theory is used instead of set theory as base theory.

The binary satisfaction predicate $\mathrm{Sat}(x, y)$ is intended to apply to formulae x and variable assignments y. The schemata of Zermelo–Fraenkel are expanded to the full language including

[13] Tarski (1936*a*) already considered axiomatic approaches. Davidson propagated an axiomatic approach in various papers (see his 1984). Halbach (2014) provides a survey of axiomatic theories of truth for arithmetical languages; Fujimoto (2012) investigates truth theories over set theory.

the satisfaction predicate. As for the axioms and rules for it, I use 'compositional' axioms that match the classical logic of the base theory.

First, $\mathrm{Sat}(x, y)$ commutes with all quantifiers and connectives. Hence we have an axiom expressing that a variable assignment satisfies a formula $A \wedge B$ if and only if it satisfies A and B; a variable assignment satisfies a formula $\neg A$ if and only if it does not satisfy A; and so on for other connectives. A variable assignment satisfies a formula $\forall x\, A$ if and only if all its x-variants satisfy A. As usual, an x-variant of a variable assignment is any variable assignment that differs from it only in the value of x. The formulae and sentences A and B may contain the satisfaction predicate.

The language of the base theory contains predicate symbols. The axioms are as expected. For instance, a variable assignment a satisfies the formula $x \in y$ if and only if $a(x) \in a(y)$, where $a(v)$ is the value of a given variable v under the variable assignment a. Similar axioms are added for all predicate symbols other than Sat. If individual constants and function symbols are present, suitable axioms have to be specified. I use the name Ω for the overall theory, comprising the axioms of the base theory, all its schemata extended to the language with Sat and the axioms for Sat.

The theory Ω resembles the usual 'Tarskian' theory of truth – with the exception that the compositional clauses are postulated also for formulae containing the satisfaction predicate. We may want to add further axioms and rules later, especially concerning the behaviour of the satisfaction predicate on formulae containing the satisfaction predicate. But for the present purposes we can proceed without additional axioms.

Since also schemata are expanded to the language with the satisfaction predicate, Ω is properly stronger than the base theory. Its consistency cannot be proved relative to Zermelo–Fraenkel set theory. However, it is equiconsistent with a typed Tarskian theory and thus there are good reasons to believe in its consistency if the base theory is assumed to be consistent. Moreover, adding analogous axioms to reasonably behaved weaker theories such as Peano arithmetic yields provably consistent extensions of these theories. The truth axioms act as a reflection principle.[14] Therefore it is at least plausible to assume the consistency of Ω.

The axioms for satisfaction describe a notion of truth based on classical logic. It reflects the axioms for classical logic in which the base theory is formulated. If a nonclassical theory were used as base theory, the axioms for justification would have to be adjusted accordingly. For instance, a theory formulated in Strong Kleene logic would require a matching theory of truth for this logic. The substitutional approach may thus be applicable to other logics. It's not a route I would like to take; but the possibility of applying the substitutional approach to nonclassical logic shows that from the substitutional approach we are unlikely to extract an argument in favour of a certain logic.

SUBSTITUTIONAL DEFINITIONS OF LOGICAL TRUTH & CONSEQUENCE

Using the axiomatized notion of truth and the defined notion of a substitutional interpretation, logical truth and consequence can now be defined. A formula A is satisfied under a

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

substitutional interpretation I and variable assignment iff the substitutional interpretation $I(A)$ of that formula is satisfied under that variable assignment. A sentence is logically valid iff it is satisfied under all substitutional interpretations and variable assignments. A sentence A follows logically from a premiss set Γ iff A is satisfied under all substitutional interpretations and all variable assignments under which all formulae in Γ are satisfied. In these definitions a formula is allowed to contain the satisfaction predicate. The definitions can be stated in the theory Ω.

A substitutional interpretation may replace an individual constants not only with an individual constant, but also with a free variable. Hence the usual worry about objects not named by any constant or another singular term is alleviated. A given constant may be replaced with a free variable and then that variable can be assigned any object via the variable assignments.

With these definitions it is trivial that logical validity implies truth. If a sentence is logically valid, it is satisfied under all substitutional interpretations and variable assignments. The identity function, that is, the function that maps every formula to itself is a substitutional interpretation. Hence, if a sentence is valid on the substitutional definition, it is satisfied under all variable assignments, that is, it is true. Similarly, it can be established that logical consequence preserves truth.

This can be explained more formally as follows. Assume that $\mathrm{Val}(x)$ is the formula in Ω defining logical truth. If $\mathrm{Val}(x)$ expresses the above notion of validity, then the claim that logical validity implies truth becomes $\forall x\,(\mathrm{Val}(x) \rightarrow \forall a\,\mathrm{Sat}(x, a))$, where $\forall a$ expresses quantification over all variable assignments (the restriction of the quantifier $\forall x$ to sentences is not necessary). This principle is not even expressible on the model-theoretic account, because there is no absolute set-theoretically definable satisfaction predicate Sat. Thus one might object that $\forall x\,(\mathrm{Val}(x) \rightarrow \forall a\,\mathrm{Sat}(x, a))$ does not express what was intended by saying that logical validity implies truth. What was really meant, so one might claim, is the schema $\mathrm{Val}([A]) \rightarrow A$ for all sentences A, where $[A]$ is some canonical name for the sentence A. This schema, however, strikes me as too weak to express the claim that logical validity implies truth. Unlike the universally quantified principle, it is not clear how to negate the schema. so we cannot even state that validity does not imply truth on the schematic account. But one can still ask whether the schema $\mathrm{Val}([A]) \rightarrow A$ is provable. In fact it is, but perhaps, the critical reader might feel, for the wrong reasons.[15]

The substitutional definitions of logical truth and consequence are universal: They apply to the entire language of Ω. In particular, it applies to sentences containing the satisfaction predicate. It shares this feature with the model-theoretic definition, because the model-theoretic definition, which is defined in set theory, applies to the entire language of set theory.

---

[15] The proof is somewhat convoluted. First it can be shown that $\mathrm{Val}([A])$ implies truth in all set-theoretic models and hence provability in (free) predicate logic by the formalized completeness theorem. The theory of satisfaction is essentially reflexive and thus proves the local reflection principle for any finite subtheory, and thus for logic. A proof via the satisfaction predicate is not feasible, even if the theory of satisfaction is (consistently) strengthened. The schema $\forall a\,\mathrm{Sat}([A], a) \rightarrow A$ is inconsistent, as Sat commutes with negation. Only the rules that allows to proceed from a proof of $\forall a\,\mathrm{Sat}([A], a)$ to A can consistently be added. (Kreisel 1965, p. 117) gives a similar argument, pointing out that it does not apply to finitely axiomatized theories.

If validity is defined in a metalanguage, as, for instance, Tarski (1936*b*) and Quine (1986) do, the universality of logic is lost. Logical consequence is then defined at best for a sublanguage of our entire language. This approach has its merits: There is nothing wrong with defining nonclassical consequence relations for some object languages in classical set theory. But this is a different project from the one pursued here; and it is different from the traditional project. We do no longer study arguments and their forms in our own language that never have true premisses and false conclusions. The definition of validity doesn't apply to our own arguments in the metalanguage.

The universality of logic is important for evaluating suggestions to tweak classical 'at the fringes'. In the theory of paradoxes it has become common to retain classical logic for the nonsemantic fragment of our language, while the full language is governed by some nonclassical language. If logic is universal, this means that classical logic has been given up and the universal laws of reasoning are no longer classical. If we define validity only for the nonsemantic part of the language, we cannot even sensibly discuss whether the notion of validity is classical, unless we go to a still higher metalanguage. We want to be able to say that a departure from classical logic for some part of the language is a departure from classical logic. Being able to do so requires a notion of validity that applies to the entire language including the semantic vocabulary.

All this does not mean that there is something wrong with these nonclassical solutions of the semantic paradoxes. It just means they have to be correctly classified as nonclassical. Intuitionists, for instance, endorse many instances of the law of excluded middle; not accepting some instances makes intuitionistic reasoning nonclassical. That classical logic applies to some sublanguage of the entire language of the intuitionist doesn't show that the notion of validity is classical.

SUBSTITUTIONAL AND MODEL-THEORETIC COUNTEREXAMPLES

The main reason for abandoning attempts to analyze logical validity via substitution has been the worry that on the substitutional account an invalid argument or an invalid sentence may declared valid.[16] In particular, there may be a model-theoretic counterexample that doesn't have a substitutional counterpart, or so it has been argued. If the language is countable, then every sentence has only countably many substitution instances. The number of model-theoretic counterexamples is not bound by any cardinality, in contrast. Hence there cannot be a different substitution instance for each model.

It can be shown that this objection does not pose a problem for the substitutional analysis, if the above substitutional definitions of validity are employed. The substitutional interpretation of a sentence can contain more free variables than the original sentence. For the sentence

[16] This historical claim needs to be substantiated. Tarski (1936*b*) rejected a version of the substitutional account characterized by his condition (F). Etchemendy discussed the persistence problem in various places in his (1990). I suspect, the reason why logicians in earlier times have not been bothered so much by the problem is that they did not fix a language and were not afraid to use demonstrative pronouns and similar devices, so that any object could be designated perhaps not by a proper name or description but at least by a pronoun.

to be valid, the substitutional interpretation must be satisfied by all variable assignments. By varying the variable assignments there is more leeway for constructing counterexamples: Counterexamples are not only obtained by substituting the nonlogical vocabulary, but also by varying the assignment of objects to the variables. On the substitutional approach advocated here, a counterexample can be understood as a pair of a substitutional interpretation and a variable assignment. Since there is no limit on the cardinality of variable assignments, there is also no limit on the cardinality of counterexamples. Hence, the above observation that there are only countably many substitution instances of a given sentence does not mean there are only countably many counterexamples.

In fact for any model-theoretic counterexample there is a substitutional counterpart, that is, a pair of a substitution instance and a variable assignment. This can be shown by proving, roughly speaking, that the claim 'A is valid in that model' can be rewritten as a substitution instance $I(A)$ of A for a suitable substitutional interpretation I. If there is a model-theoretic counterexample M to A, that is, $M \nvDash A$, then the pair with I and an assignment a that assigns M to 'that model' is a substitutional counterexample. Thus any model-theoretic counterexample can be converted into a substitutional counterexample. It follows that, if a sentence is logically valid on the substitutional account, it is also valid on the usual model-theoretic account. The reasoning can be carried out in the theory Ω and stated more precisely as follows:

> SUBSTITUTIONAL PLENITUDE In Ω the following statement is provable: There is a substitutional interpretation I such that for all models M, assignments a, and sentences A, $M \vDash \phi$ iff $\mathrm{Sat}(I(A), a(m/x))$. Here $a(m/x)$ is the result of changing (if needed) the assignment a so that the object m is assigned to the variable x. Here x is some fixed variable, for instance the first in the alphabet.

The idea of the proof is simple: The claim $M \vDash A$ for models M and sentences A is formulated in set theory. It is shown that $m \vDash A$ with m as a free variable is equivalent to a substitution instance of A by pushing in $M \vDash$ through all connectives and quantifiers of A, relativizing quantifiers of A to the domain of m along the way. The resulting formula is satisfied by a variable assignment a that assigns M to m iff $M \vDash A$. This is proved in Ω, using the axioms for the satisfaction predicate.

Proof. I outline how to prove the result in Ω. Let A be a formula and $x_1, \ldots, x_n$ be the finite list of all variables occurring in A whether they are occur free or not. The substitutional interpretation I is now defined in the following way. An atomic subformula in A can be of the form Rx_1x_2. The interpretation I maps Rx_1x_2 to the formula $m \vDash Rx_1x_2(a(x_1/[x_1], x_2/[x_2]))$. The latter formula is still formulated in the language of set theory, and thus in the language of the base theory; it expresses that the formula Rx_1x_1 is holds in the model m under the variable assignment a but with the value for the variable 'x_1' changed into x_1 and the value for the variable 'x_2' changed into x_2. This move allows one to bind the variables 'x_1' and 'x_2' from the 'outside'. The formula has m, a, x_1, and x_2 as free variables. In fact, the formula A itself could be quantified away. The predicate symbol R could be the satisfaction predicate; but then the overall formula is still in the language of set theory, because R is only mentioned and not used. Other atomic formulae are dealt with in an analogous way. The interpretation restricts all quantifiers in A to the domain of m. This requires a function d which yields applied to a model its domain. Thus the formula $I(\forall x\, B)$ is $\forall x\, (d(m) \rightarrow I(B))$. The substitutional interpretation $I(A)$ of the sentence A will contain two free variables m and a. Validity in m commutes with all connectives and quantifiers in A, because all quantifiers are restricted to the domain of m. Thus $I(A)$ is equivalent in set theory (and thus Ω) to $m \vDash A(a)$. Now let b be a variable assignment that assigns M to m and some object to a (the value of a would only matter, if A contained free variables). Then $m \vDash A(a)$ and $\mathrm{Sat}([m \vDash A(a)], b)$ are equivalent in Ω. This concludes the proof.

The proof can be carried out in various settings by tweaking it slightly. In particular, it applies to the variations of the definition of substitutional interpretations and the related free logics. Moreover, the proof doesn't depend on set theory as the base theory for Ω. However, it's the most interesting case because the argument above shows that all the model-theoretic counterexamples conceived as sets can be converted into substitutional counterexamples. Model theory can be developed also in much weaker theories, notably in arithmetic. Of course in arithmetic we are confined to models that can somehow be coded in the numbers. The theory Ω can then be reformulated with arithmetic, say Peano arithmetic, as base theory. The formulation of the axioms for satisfaction becomes a little more cumbersome as variable assignments have to be finite functions. But the argument for substitutional plenitude will still go through. I am mentioning this possibility only because it shows that the argument doesn't rely on the specifics of set theory and is fairly stable with respect to variations of the base theory.

The proof relies on the definition of substitutional interpretations as functions that can introduce new free variables. The usual worry about persistence is often stated by saying that there may be objects for which we do not have names. Thus, on the substitutional account, they cannot be used for constructing counterexamples. This worry is alleviated by allowing free variables in the substitution instances that can be assigned any object by a suitable variable assignment. Then one may also worry that there could be relations that cannot be expressed by a suitable formula in the language. Again the persistence proof shows that once we can express $\vDash$ in the language and allow free variables in the substitution instances any set-theoretic model can be transformed into a 'substitutional model', that is, a pair of a substitutional interpretation and a variable assignment. Hence a substitutional definition of validity, as given above, is not very sensitive to the choice of the nonlogical vocabulary. Just with very little vocabulary, in particular with only $\in$, we can show that any set-theoretic model has a substitutional counterpart and the worry about the inexpressibility of certain relations does not arise.

THE FORMALLY RIGOROUS SQUEEZING ARGUMENT

The proof of substitutional plenitude can be used to rephrase Kreisel's squeezing argument with substitutional validity slotted into the place of Kreisel's elusive informal and intuitive validity. The notion of substitutional validity is not purely set-theoretic, but nevertheless formally rigorous. Thus a formally rigorous version of the squeezing argument is obtained:

$$\begin{tikzcd} \vdash_{\mathrm{ND}} \phi \arrow[r, "\text{soundness}", "\text{theorem}"'] & \text{ϕ is substitutionally valid} \arrow[r, "\text{substitutional}", "\text{plenitude}"'] & \vDash \phi \arrow[ll, bend left, "\text{completeness theorem}"] \end{tikzcd}$$

The proof that provability in Natural Deduction implies validity can be formalized and proved in Ω in a formally rigorous way. It is a simple inductive proof that makes the informal argument in Kreisel's original version of the proof explicit and formal. To establish that substitutional validity implies validity in all set-theoretic models, the proof of substitutional plenitude in the preceding section is used. Again this is a proof in the formal theory Ω. Hence all implications in the diagram are formally rigorous.

It is not obvious that model-theoretic validity, that is $\vDash \phi$, implies substitutional validity. There are substitutional countermodels that have no direct set-theoretic counterpart. In particular, if a substitution instance does not have quantifiers relativized to a formula defining a set, then it is at least not clear what the corresponding set-theoretic model could be. Unsurprisingly in this respect substitutional validity behaves like untuitive validity: Both intuitive and substitutional validity can be understood as notions of validity that conform with the Generalized Tarski Thesis; both do not rule out the intended of trivial interpretation as a case (in the sense of the Generalized Tarski Thesis), while the model-theoretic account rules out this interpretation as a case.[17]

The squeezing argument still serves its purpose: It shows that by concentrating on set-theoretic countermodels the set of valid sentences doesn't change. In its formally rigorous form the squeezing argument still shows that the model-theoretic definition succeeds in classifying sentences and arguments correctly as logically valid or not, even though we lack a set-theoretic 'intended' model and other 'class sized' models.

Using the squeezing argument also the three different definitions of logical consequence – the proof-theoretic, the substitutional, and the model-theoretic – can be shown to declare the same arguments valid.[18] It follows that logical consequence in the substitutional sense advocated here is compact. This is in contrast to Quine's (1986) version of the substitutional definition of logical validity, which is not compact as Boolos (1975) observed

Certain other conceptions of substitutional validity don't fit into Kreisel's squeezing argument in the same way as the notion of substitutional validity defended here. I have defined validity using a primitive type-free truth predicate. It could be asked what happens if the axioms for the satisfaction predicate are restricted so that the satisfaction predicate applies

[17] Participants of the FilMat16 in Chieti suggested the use of set-theoretic reflection principles to show a direct equivalence of substitutional and model-theoretic validity without going through the Gödel completeness theorem. This strategy has its limitation if the equivalence is to be shown for arguments with infinite premiss sets.

[18] Again the account can be applied to variations based on different version of free logic, for instance. In this case the systems of Natural Deduction and the model-theoretic definitions must be adjusted, of course.

only to formulae without the satisfaction predicate. For instance, we would only postulate that satisfaction commutes with conjunction only for the formulae of the base language, that is, for formulae without Sat. In this case only substitution instances without the satisfaction predicate would be admitted. The set of valid sentences that are classified as valid wouldn't change. Basically we only would have formalized truth by a primitive new predicate in the language of Ω that is usually defined in a metatheory or left as an informal metatheoretic notion as in (Quine 1986) or in Tarski's (1936*b*) substitutional definition (F) of validity (which he rejected). In fact, substitutional plenitude and thus the squeezing argument could be proved in the same way as before, thereby solving the problem of overgeneration of the substitutional approach to validity. So why not adopt a safer typed notion of satisfaction? We would no longer quantify over all interpretations. In particular, the trivial intended interpretation of sentences with the satisfaction predicate would no longer be among the substitution instances. The satisfaction predicate would have to be replaced with a formula not containing the satisfaction predicate. Technically it wouldn't make a change because substitution instances without the satisfaction predicate suffice for producing counterexamples for validity in the same way set-sized models suffice on the model-theoretic account. But the definition of validity could no longer claim to be a direct analysis of the notion of validity. It would need an argument – presumably an appeal to the completeness theorem – to show that it does its job and defines the correct extension of validity. If that is the only aim, a much leaner definition of substitutional validity could be used that does not even require a type truth predicate.[19]

To obtain a full conceptual analysis of logical truth and consequence we need to quantify over all substitutional interpretations. The class of substitution instances must not be restricted, even if it can be shown that a particular restriction does not change the extension of the definition. The definition with unrestricted quantification over substitutional interpretations, however, requires a primitive type-free truth predicate.

[19] If the only concern is to give a extensionally correct substitutional definition of validity, we can endorse set-theoretic reductionism and define validity substitutionally without a primitive truth or satisfaction predicate. The models constructed in the proof of the Henkin completeness theorem for first-order logic are basically maximal consistent sets of sentences. As Kleene (1952, chapter XIV) observed, we can limit ourselves to maximal consistent Δ^0_2-sets. Hence validity is equivalent to membership in all these Δ^0_2-sets. Thus a sentence is valid if and only if it satisfies all the defining formula of these Δ^0_2-sets. The satisfaction relation is definable in set theory as a partial satisfaction predicate for Δ^0_2-formulae. Now the defining Δ^0_2-formulae can be pushed into the sentence in the style of the proof of substitutional plenitude, so that the membership in one of these Δ^0_2-sets is actually expressed by a substitution instance in the sense defined above. The relativization of the quantifiers to the set of Henkin constants can be handled via free variables. This shows that there are extensionally correct substitutional definitions of validity in set theory without any additional semantic notions. However, such definitions suffer from the same problem as the usual model-theoretic definitions of validity: They are not direct analyses of logical validity. That they capture the intuitive notion of validity and are indeed equivalent to the substitutional definition of validity given above can be seen only by some sophisticated arguments.

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

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

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

Bolzano, Bernhard (1837), *Wissenschaftslehre. Versuch einer ausführlichen und grösstentheils neuen Darstellung der Logik mit steter Rücksicht auf deren bisherige Bearbeiter*, J. E. v. Seidel, Sulzbach. four volumes.

Boolos, George (1975), 'On second-order logic', *Journal of Philosophy* 72, 509–527.

Davidson, Donald (1973), In defence of convention T, *in* H.Leblanc, ed., 'Truth, Syntax and Modality', North Holland Publishing Company, pp. 17–36. Reprinted in Davidson (1984).

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

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

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

Etchemendy, John (2008), Reflections on consequence, *in* D.Patterson, ed., 'New Essays on Tarski and Philosophy', Oxford University Press, Oxford, pp. 263–299.

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

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

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

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

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

Kleene, Stephen Cole (1952), *Introduction to Metamathematics*, North Holland.

Kreisel, Georg (1965), Mathematical logic, Vol. III of *Lectures on Modern Mathematics*, 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.

Kripke, Saul (1972), Naming and Necessity, *in* D.Davidson and G.Harman, eds, 'The Semantics of Natural Languages', Reidel, Dordrecht, pp. 253–355.

McGee, Vann (1992), 'Two Problems with Tarski's Theory of Consequence', *Proceedings of the Aristotelian Society* 92(92), 273–292.

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

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

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

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*a*), 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica Commentarii Societatis Philosophicae Polonorum* 1.

Tarski, Alfred (1936*b*), 'Über den Begriff der logischen Folgerung', *Actes du congrès international de philosophie scientifique* 7, 1–11. translation "On the Concept of Logical Consequence" in (Tarski 1983, 409–20).

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

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

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

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

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