

The Substitutional Theory of Logical Consequence

Volker Halbach

B.Phil. Seminar Logic and Philosophy of Logic

11th February 2019

I aim at a *semantic* theory of logical consequence.

Example

The following argument is not valid:

Some cats are animals. Therefore all cats are animals.

Counterexample:

*Some cats are **black**. Therefore all cats are **black**.*

An argument is valid iff there is no interpretation of the nonlogical vocabulary under which all premisses are true and the conclusion false.

Cf. ‘Tarski’s Generalized Thesis’ in Beall and Restall (2006) and earlier.

The notion of logical consequence should be formal, **universal** and **express truth preservation under *all* interpretations, including the intended interpretation.**

- ▶ Logic is *universal*. In particular, it applies to my entire language, in which logical consequence is defined.
- ▶ Logical consequence is *truth preservation under all interpretations* of the nonlogical vocabulary. We must not omit any interpretation, especially not the ‘intended’ interpretation.

Theories of logical consequence differ mainly in what counts as an interpretation. The required notion of truth depends on what interpretations are (and vice versa).

The notion of logical consequence should be formal, **universal** and **express truth preservation under *all* interpretations, including the intended interpretation.**

- ▶ Logic is *universal*. In particular, it applies to my entire language, in which logical consequence is defined.
- ▶ Logical consequence is *truth preservation under all interpretations* of the nonlogical vocabulary. We must not omit any interpretation, especially not the ‘intended’ interpretation.

Theories of logical consequence differ mainly in what counts as an interpretation. The required notion of truth depends on what interpretations are (and vice versa).

Why is *truth preservation under **all** interpretations* important?
The ‘intended model’ of set theory (if there is one) is not among the models in the sense of model theory. Should we worry that the following is possible?

- ▶ ϕ isn't a logical consequence of Γ .
- ▶ for all models M (in the sense of model theory): if $M \models \gamma$ for all $\gamma \in \Gamma$, then $M \models \phi$

Assumption: We work in a standard one-sorted first-order language expanding the language of set theory without class quantifiers or the like (but we may have relativizing predicates for sets and urelements).

Logical constants are $\neg$, $\wedge$, and $\forall$ (later also $=$).

I don't consider higher-order quantifiers, cardinality quantifiers, modal operators, etc.

Why is *truth preservation under **all** interpretations* important? The ‘intended model’ of set theory (if there is one) is not among the models in the sense of model theory. Should we worry that the following is possible?

- ▶ ϕ isn't a logical consequence of Γ .
- ▶ for all models M (in the sense of model theory): if $M \models \gamma$ for all $\gamma \in \Gamma$, then $M \models \phi$

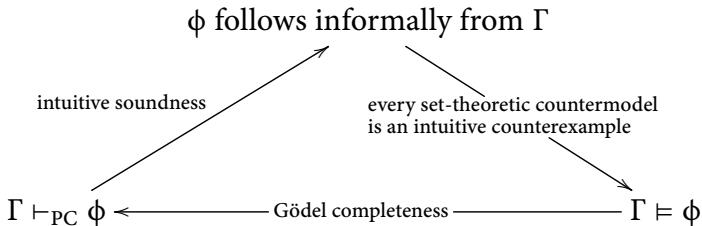
Assumption: We work in a standard one-sorted first-order language expanding the language of set theory without class quantifiers or the like (but we may have relativizing predicates for sets and urelements).

Logical constants are $\neg$, $\wedge$, and $\forall$ (later also $=$).

I don't consider higher-order quantifiers, cardinality quantifiers, modal operators, etc.

Of course we don't have to worry because of the Gödel Completeness theorem.

Kreisel (1965, 1967a,b) argued for the extensional correctness of the model-theoretic definition for first-order logic with his *squeezing argument*:



$\Gamma \models \phi$ means that ϕ follows model-theoretically from Γ ;
 $\Gamma \vdash_{PC} \phi$ means that ϕ is provable from Γ (in, say, Natural Deduction).

For the Completeness theorem and the Squeezing argument we don't have to quantify over *all* set-theoretic models. We could equally use the following definitions:

$\Gamma \models \phi$ iff (for all **countable** models M : if $M \models \gamma$ for all $\gamma \in \Gamma$, then $M \models \phi$).

$\Gamma \models \phi$ iff (for all **Δ_2^0 -definable** models M : if $M \models \gamma$ for all $\gamma \in \Gamma$, then $M \models \phi$).

These analyses are not adequate analyses, although they are extensionally adequate.

The Completeness theorem shows the extensional adequacy of the model-theoretic definition, but we don't have an analysis of what Kreisel calls 'informal validity'.

Interpretations have been defined as

- ▶ set-theoretic models · model theory since Tarski and Vaught (1956)
- ▶ mental objects · Bolzano (1837)
- ▶ higher-order objects (assignments) · Tarski (1936), Williamson (2000)
- ▶ syntactic substitutions · Buridan, Quine (1986), Eder (2016)

Corresponding to the notion of interpretation, a notion of truth relative to such an interpretation is required.

	logic is universal	intended interpretation	no primitive sem. notions
model theory	yes	no	yes
Tarski (1936)	no	yes	yes
Quine (1986)	yes	no	yes
VH	yes	yes	no

The entries may need some qualifications.

The first three theories are reductionist in the sense that all semantic notions employed are eliminated and defined in purely non-semantic terms (again with some qualifications).

My theory is substitutional in the sense that interpretations are understood substitutionally. But substitution instances do not come only from some restricted language such as the language of arithmetic.

Substitutional validity is defined as follows:

A sentence is logically valid iff all its substitution instances are true.

An argument with premisses in Γ and conclusion ϕ is valid iff for all substitutional interpretations: if all substitution instances of sentences in Γ are true, then the substitution instance of ϕ is true.

Example

Some cats are animals. Therefore all cats are animals.

Substitutional counterexample:

*Some cats are **black**. Therefore all cats are **black**.*

Substitutional validity is defined as follows:

A sentence is logically valid iff all its substitution instances are true.

An argument with premisses in Γ and conclusion ϕ is valid iff for all substitutional interpretations: if all substitution instances of sentences in Γ are true, then the substitution instance of ϕ is true.

Example

Some cats are animals. Therefore all cats are animals.

Substitutional counterexample:

*Some cats are **black**. Therefore all cats are **black**.*

Precursors

Buridan defines such a notion, but doesn't call it logical consequence.

Bolzano (1837) is often cited as a proponent of the substitutional definition, e.g., by Etchemendy (1990); but that's probably incorrect.

Quine (1986) built on the arithmetized Completeness theorem. See Eder (2016).

General worry: The substitutional account is worse than the model-theoretic analysis, because in the definition of logical consequence we quantify only over substitution instances. This was exactly the reason why Tarski (1936) rejected it and developed a precursor of the model-theoretic account.

My aim is not only to obtain an extensionally correct substitutional definition of logical consequence, but an (intensionally) adequate definition.

Substitutional interpretations are functions that uniformly replace the nonlogical vocabulary in all formulae of the language with suitable expressions.

I focus on the binary predicate symbols $\in$ and Sat as nonlogical symbols; other predicate and function symbols, including constants, can be dealt with by the usual methods. Identity can be treated as logical constant or not.

I allow free variables in the substitution instances. This will overcome the problems of other substitutional accounts.

Substitutional interpretations are functions that uniformly replace the nonlogical vocabulary in all formulae of the language with suitable expressions.

I focus on the binary predicate symbols $\in$ and Sat as nonlogical symbols; other predicate and function symbols, including constants, can be dealt with by the usual methods. Identity can be treated as logical constant or not.

I allow free variables in the substitution instances. This will overcome the problems of other substitutional accounts.

A function I from the set of all formulae Form_{Sat} into Form_{Sat} is a substitutional interpretation iff there are formulae $\sigma_{\in}(x, y)$, $\sigma_{=}(x, y)$, $\sigma_{\text{Sat}}(x, y)$ and possibly $\delta(x)$ such that the following two conditions hold:

If ϕ has been obtained by the substitution above, I has the following property:

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

In the last line $\forall x (\delta(x) \rightarrow I(\psi))$ is meant if there is a $\delta(x)$; otherwise it's $\forall x \psi(x)$.

Some renaming of variables is required. I suppress this.

Assume we have a binary symbols R, S and unary symbols P and Q in the language.

$\forall x (Px \rightarrow \exists y Ryx)$ (original formula)

$\forall x (Qxv \rightarrow \exists y \neg \exists z (Qyv \wedge Sxz))$ (substitution instance)

$\forall x (\underline{Rzx} \rightarrow (\neg Sxx \rightarrow \exists y (\underline{Rzy} \wedge \neg Syx)))$ (subst. inst.)

The underlined formula is the relativizing formula. z is an additional parameter.

original argument

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

substitution instance

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

substitution instance

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

The axioms for Sat are added to our overall theory, an extension of ZF. Syntax is coded as usual. Schemata of the base theory are extended to the language with Sat (and possibly D). The quantifiers for a and b range over variables assignments, the quantifiers for ϕ and ψ over formulae of the *entire* language (including Sat and D).

The following axioms are obligatory:

$$\forall a \forall v \forall w \left(\text{Sat}(\ulcorner Rvw \urcorner, a) \leftrightarrow Ra(v)a(w) \right)$$

and similarly for all predicate symbols other than Sat

$$\forall a \forall \phi \left(\text{Sat}(\ulcorner \neg \phi \urcorner, a) \leftrightarrow \neg \text{Sat}(\ulcorner \phi \urcorner, a) \right)$$

$$\forall a \forall \phi \forall \psi \left(\text{Sat}(\ulcorner \phi \wedge \psi \urcorner, a) \leftrightarrow (\text{Sat}(\ulcorner \phi \urcorner, a) \wedge \text{Sat}(\ulcorner \psi \urcorner, a)) \right)$$

$$\forall a \forall v \forall \phi \left(\text{Sat}(\ulcorner \forall v \phi \urcorner, a) \leftrightarrow \forall b \left('b \text{ is } v\text{-variant of } a' \rightarrow (\text{Sat}(\ulcorner \phi \urcorner, b)) \right) \right)$$

extensionality of Sat.

Optional axioms:

Determinateness axioms:

$$\forall a \forall v \forall w \left(D(\ulcorner Rvw \urcorner, a) \right)$$

and similarly for predicate symbols other than Sat, but including D

$$\forall a \forall \phi \left(D(\ulcorner \neg \phi \urcorner, a) \leftrightarrow D(\ulcorner \phi \urcorner, a) \right)$$

$$\forall a \forall \phi \forall \psi \left(D(\ulcorner \phi \wedge \psi \urcorner, a) \leftrightarrow (D(\ulcorner \phi \urcorner, a) \wedge D(\ulcorner \psi \urcorner, a)) \right)$$

$$\forall a \forall v \forall \phi \left(D(\ulcorner \forall v \phi \urcorner, a) \leftrightarrow \forall b \left('b \text{ is } v\text{-variant of } a' \rightarrow (D(\ulcorner \phi \urcorner, b)) \right) \right)$$

Sat-iteration axiom:

$$\forall a \forall \phi \forall v \left(D(\ulcorner \phi \urcorner, a(v)) \rightarrow \left(\text{Sat}(\ulcorner \text{Sat}(\ulcorner \phi \urcorner, v) \urcorner, a) \leftrightarrow \text{Sat}(\ulcorner \phi \urcorner, a(v)) \right) \right)$$

Call the theory with Sat axioms CD.

'A sentence is logically valid iff all its substitution instances are true.'

substitutional definition of logical truth in CD

$$\forall \phi \left(\text{Val}(\ulcorner \phi \urcorner) :\leftrightarrow \forall I \forall a \text{Sat}(I\ulcorner \phi \urcorner, a) \right)$$

Here $\forall I$ ranges over substitutional interpretations, as defined above.

Similarly, an argument is logically valid iff there is no substitutional interpretation and no variable assignment that make the premisses true and the conclusion false.

substitutional definition of logical validity in CD

$$\Gamma \models_s \phi \text{ iff } \forall I \forall a \left(\left(\forall \gamma \in \Gamma \text{ Sat}(I(\gamma), a) \right) \rightarrow \text{Sat}(I(\phi), a) \right)$$

On this definition logic is *universal*, just as under the model-theoretic definition.

A typed theory of Sat such as T(ZF) instead of CD would suffice for this.

Similarly, an argument is logically valid iff there is no substitutional interpretation and no variable assignment that make the premisses true and the conclusion false.

substitutional definition of logical validity in CD

$$\Gamma \models_s \phi \text{ iff } \forall I \forall a \left(\left(\forall \gamma \in \Gamma \text{ Sat}(I(\gamma), a) \right) \rightarrow \text{Sat}(I(\phi), a) \right)$$

On this definition logic is *universal*, just as under the model-theoretic definition.

A typed theory of Sat such as T(ZF) instead of CD would suffice for this.

$\models_S$ is consequence in free logic. This is because nothing rules out the use of $x \notin x \wedge x \in x$ as relativizing formula $\delta(x)$. In (Halbach, 2018) I added $\exists x \delta(x)$ to the premisses to get full classical logic.

But in a way I am happy with the free logic. On the model-theoretic account it's ruled out for practical reasons only.

We can also drop relativizing formulae and get constant domain semantics like Tarski (1936) and Williamson (2000). Then, e.g. $\exists x \exists y x \neq y$ comes out as logical truth, if identity is treated as logical constant.

Thus, modulo the tweaks above, $\models_S$ is equivalent to model-theoretic logical consequence. This will be proved below.

$\models_S$ is consequence in free logic. This is because nothing rules out the use of $x \notin x \wedge x \in x$ as relativizing formula $\delta(x)$. In (Halbach, 2018) I added $\exists x \delta(x)$ to the premisses to get full classical logic.

But in a way I am happy with the free logic. On the model-theoretic account it's ruled out for practical reasons only.

We can also drop relativizing formulae and get constant domain semantics like Tarski (1936) and Williamson (2000). Then, e.g. $\exists x \exists y x \neq y$ comes out as logical truth, if identity is treated as logical constant.

Thus, modulo the tweaks above, $\models_S$ is equivalent to model-theoretic logical consequence. This will be proved below.

$\models_S$ is consequence in free logic. This is because nothing rules out the use of $x \notin x \wedge x \in x$ as relativizing formula $\delta(x)$. In (Halbach, 2018) I added $\exists x \delta(x)$ to the premisses to get full classical logic.

But in a way I am happy with the free logic. On the model-theoretic account it's ruled out for practical reasons only.

We can also drop relativizing formulae and get constant domain semantics like Tarski (1936) and Williamson (2000). Then, e.g. $\exists x \exists y x \neq y$ comes out as logical truth, if identity is treated as logical constant.

Thus, modulo the tweaks above, $\models_S$ is equivalent to model-theoretic logical consequence. This will be proved below.

$\models_S$ is consequence in free logic. This is because nothing rules out the use of $x \notin x \wedge x \in x$ as relativizing formula $\delta(x)$. In (Halbach, 2018) I added $\exists x \delta(x)$ to the premisses to get full classical logic.

But in a way I am happy with the free logic. On the model-theoretic account it's ruled out for practical reasons only.

We can also drop relativizing formulae and get constant domain semantics like Tarski (1936) and Williamson (2000). Then, e.g. $\exists x \exists y x \neq y$ comes out as logical truth, if identity is treated as logical constant.

Thus, modulo the tweaks above, $\models_S$ is equivalent to model-theoretic logical consequence. This will be proved below.

The identity function with $I(\phi) = \phi$ on sentences is a substitutional interpretation. Thus we have:

- ▶ Logical validity (logical truth) trivially implies truth, i.e., $CD \vdash \forall x (\text{Val}(x) \rightarrow \forall a \text{Sat}(x, a))$.
- ▶ Similarly, logical consequence preserves truth.

On the substitutional account, the ‘intended’ interpretation is completely trivial. There is no need for mysterious intended models (whose existence is refutable), indefinite extensibility, or the impossibility of unrestricted quantification.

This requires an untyped theory of Sat , because we admit substitution instances with Sat .

Logical consequence is truth preservation under all interpretations of the nonlogical vocabulary. We must not omit any interpretation, especially not the ‘intended’ interpretation.

We do have the intended interpretation (and other ‘class-sized’ interpretations); but what about all the set-theoretic models?

Let $\langle D, E, S \rangle$ be a model with domain D , extension E for $\in$ and S for Sat . Variables may have to be renamed.

Lemma

$$\text{CD} \vdash \forall \phi \forall a \in D^\omega \left(\langle D, E, S \rangle \models \phi[a] \leftrightarrow \text{Sat}(I_1(\phi), a_{\mathcal{M}}^{I_1}) \right)$$

$$I_1(\phi) := \begin{cases} \langle x, y \rangle \in v_2 & \text{if } \phi \text{ is } x \in y, \\ x = y & \text{if } \phi \text{ is } x = y, \\ \langle x, y \rangle \in v_3 & \text{if } \phi \text{ is } \text{Sat}(x, y), \\ \neg I_1(\psi) & \text{if } \phi \text{ is } \neg \psi, \\ I_1(\psi) \wedge I_1(\chi) & \text{if } \phi \text{ is } \psi \wedge \chi, \text{ and} \\ \forall x (x \in v_1 \rightarrow I_1(\psi)) & \text{if } \phi \text{ is } \forall x \psi \end{cases}$$

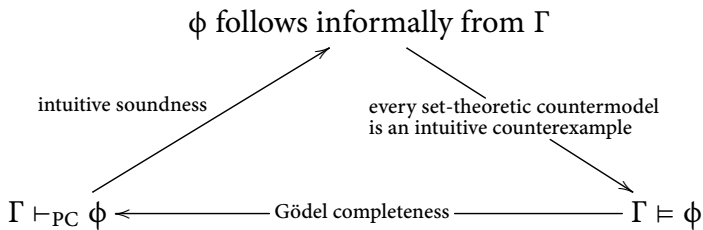
$$a_{\mathcal{M}}^{I_1}(v_1) := D$$

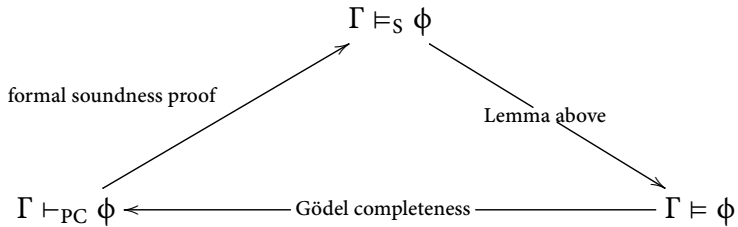
$$a_{\mathcal{M}}^{I_1}(v_2) := E$$

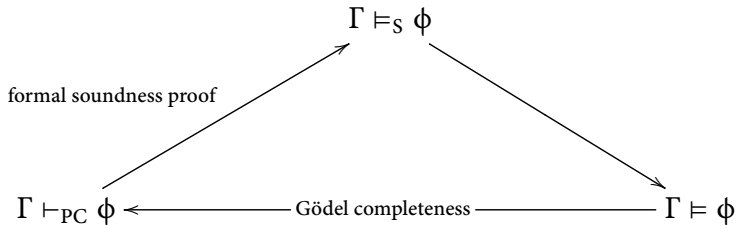
$$a_{\mathcal{M}}^{I_1}(v_3) := S$$

$$a_{\mathcal{M}}^{I_1}(v_{n+3}) := a(v_n)$$

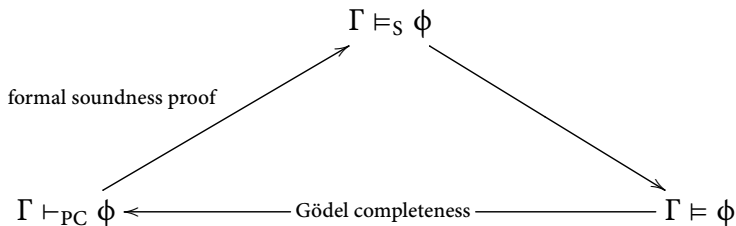
A substitutional model is a pair $\langle I, a \rangle$ of a substitutional interpretation and a variable assignment. Every set-theoretic model corresponds to the substitutional model $\langle I_1, a_{\mathcal{M}}^{I_1} \rangle$. But not every substitutional model corresponds to a set-theoretic model (for instance, if there is no relativizing formula $\delta(x)$ or $\delta(x)$ doesn't define a set).







The squeezing argument is now a reduction of a semantic notion defined from an irreducible semantic satisfaction predicate to a purely mathematical notion.



The squeezing argument is now a reduction of a semantic notion defined from an irreducible semantic satisfaction predicate to a purely mathematical notion.

The squeezing argument for substitutional validity naturally slots into the place of Kreisel's (1965; 1967a) 'informal logical validity'. There is no need to postulate an elusive informal notion of logical validity.

In particular, we have:

THEOREM

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

The substitutional notion of validity isn't very sensitive to the choice of the signature, if free variables are admitted in the substitution instances and the base theory is sufficiently strong.

The substitutional notion of validity isn't very sensitive to the choice of Sat-axioms.

In particular, we have:

THEOREM

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

The substitutional notion of validity isn't very sensitive to the choice of the signature, if free variables are admitted in the substitution instances and the base theory is sufficiently strong.

The substitutional notion of validity isn't very sensitive to the choice of Sat-axioms.

On my substitutional account

- ▶ **Logic is universal.** There is no restriction to a weaker object language. This requires a type-free truth predicate.
- ▶ **Logical consequence is truth preservation under all interpretations.**
 1. The intended interpretation and other ‘class sized’ models are among the interpretations we quantify over in the definition of validity.
 2. Every set-theoretic model corresponds to the substitutional model $\langle I_1, a_{\mathcal{M}}^{I_1} \rangle$.

For discussion:

(Informal) logical validity is substitutional validity.

Worries:

1. Can there be further interpretations we don't quantify over?
2. Shouldn't logical consequence be defined in purely mathematical or, at least, nonsemantic terms?
3. We shouldn't need a fancy type-free truth or satisfaction predicate to define logical consequence.

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

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

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

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

Volker Halbach. The substitutional analysis of logical consequence. *Noûs*, 2018. to appear.

Georg Kreisel. Mathematical logic. In T. L. Saaty, editor, *Lectures on Modern Mathematics III*, pages 95–195. Wiley, New York, 1965.

Georg Kreisel. Informal rigour and completeness proofs. In Imre Lakatos, editor, *The Philosophy of Mathematics*, pages 138–171. North Holland, Amsterdam, 1967a.

Georg Kreisel. Mathematical logic: what has it done for the philosophy of mathematics? In Ralph Schoenman, editor, *Bertrand Russell: Philosopher of the Century*, pages 201–272. George Allen & Unwin, London, 1967b.

Willard V. Quine. *Philosophy of Logic*. Harvard University Press, second edition, 1986.

Alfred Tarski. Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique*, volume 7, Paris, 1936.

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

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