Logical Constants and Unrestricted Quantification

Volker Halbach

Abstract: Variants of the so-called permutation criterion have been used for distinguishing between logical and non-logical operations or expressions. Roughly, an operation is defined as logical if, and only if, it is invariant under arbitrary permutations on every domain. Thus a logical operation behaves on all objects in the same way. An expression is logical if, and only if, the operation expressed by it is logical. I consider a variant of the permutation criterion that eliminates domains: An operation is permutation-invariant if, and only if, it is invariant under arbitrary permutations of the universe. An expression is logical if, and only if, it expresses an operation that is permutation-invariant in this sense. This domain-free definition of the invariance criterion matches definitions of logical consequence without domains where first-order quantifiers are taken to range over all (first-order) objects in all interpretations. Without domains some problems of the invariance criterion disappear. In particular, an operation can behave on all objects of any domain in the same way, while still behaving very differently in each domain. On the criterion without domains, a logical operation always behaves on all objects in the same way, not only on all objects of any given domain.

Keywords: Logical consequence, Logical constants, Permutation criteria for logical constants, Absolutely unrestricted quantification, Second-order logic

An argument is logically valid if, and only if, there is no interpretation of the non-logical vocabulary under which the premisses are all true and the conclusion is false. On this semantic definition, logical consequence depends on the distinction between the logical and non-logical vocabulary. The logical vocabulary is widely taken to be the vocabulary that is independent of the subject matter and behaves on all objects in the same way. Logicians have tried to cast this characterization into a more formal criterion: An expression is a logical constant if, and only if, the behaviour of the expression cannot be changed by replacing every object with a proxy in such a way that every object is a proxy of something. In other words, the behaviour of logical constants is invariant under arbitrary permutations.

Logicians have discussed various versions of this invariance criterion. Usually, the behaviour of a term or the operations expressed by the term is said to be logical if, and only if, the behaviour of the term or operation, when interpreted in a model with a certain domain, is not affected by permuting the model's domain. This yields a criterion relative to a model; but one can usually choose the 'intended' interpretation over the domain. This model-theoretic version of the invariance criterion is formulated in set theory, even if it is applied to type-theoretic languages.

The model-theoretic criterion of logicality deviates from the initial informal characterization as follows: According to the formal criterion, a term or operation is defined as logical if, and only if, it behaves on all objects of any domain in the same way, while the original informal characterization takes a term or operation to be logical if, and only if it behaves on all objects in the same way; no mention is made of domains. In this paper I discuss a formal version of the invariance criterion without domains. It is closer to the original conception of

Volker Halbach, University of Oxford, UK, volker.halbach@philosophy.ox.ac.uk, ORCID: 0000-0001-7628-1118.

Journal for the Philosophy of Mathematics, Vol. 1, 99-112, 2024 | ISSN 3035-1863 | DOI: 10.36253/jpm-2935

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.
Competing Interests: The Author(s) declare(s) no conflict of interest.

Firenze University Press | https://riviste.fupress.net/index.php/jpm

independence from the subject matter than the model-theoretic. Not only is my criterion closer to the initial informal characterization of logical constants; it also avoids certain well-known problems that are generated by the use of domains.

Of course, domains are essential to the understanding of interpretations as (set-theoretic) models and, consequently, to the model-theoretic definition of logical consequence. However, there are also definitions of logical consequence that do not rely on domains. For such definitions my invariance criterion is much better suited than the model-theoretic. Before delving into the discussion of invariance criteria, I explain why one might want to eliminate domains from the definition of logical validity.

1. Two strategies for defining logical consequence

Two strategies for defining logical truth and consequence have their origin in Tarski's work. One of them features domains; the other does without. The latter strategy – let us call it *the type-theoretic strategy* – is associated with his brief conference paper (1936), which was an application of his theory of truth developed in *The Concept of Truth in the Formalized Languages* (1935). On this strategy, a formula is classified as logically valid if, and only if, the result of replacing all non-logical terms in it with variables of the appropriate type is satisfied by all variable assignments. For instance, the sentence $\forall x\,(Px \to Qx \vee Px)$ is declared logically valid, because $\forall x\,(Xx \to Yx \vee Xx)$ is satisfied by all (second-order) variable assignments. The techniques developed by Tarski in *The Concept of Truth* yield a definition of satisfaction of a formula of order n in a language of order $n+1$ relative to a variable assignment. No mention is made of domains of quantification on Tarski's type-theoretic strategy.[1]

The second strategy devised by Tarski is the usual model-theoretic definition of logical consequence. There is no need to ascend to higher-order languages on this approach: Logical truth and consequence for first-order sentences can be defined in first-order set theory. Tarski developed this strategy only later, after first-order set theory had largely replaced type-theory as the foundations of mathematics.[2] The domain of quantification in a model is always a set, at least if we work in first-order set theory in the usual way. Satisfaction over a class-sized domain cannot be generally defined within first-order set theory. Logical truth or validity of a sentence is then defined as truth, that is, satisfaction by all variable assignments, in all models.

Logical consequence, that is, the logical validity of an argument, can be defined in a way that is analogous to that of logical truth or validity of a single sentence on both, the model- and the type-theoretic strategy.

Here in this paper I focus on definitions of logical consequence for first-order languages with a special interest in the language $\mathscr{L}_{\mathrm{ZF}}$ of set theory and its finite first-order expansions. I see these languages as regimented versions of the language that many logicians use for theorizing; they are the languages in which logical consequence can be defined in the model-theoretic way. No ascent to a higher-order language is required. On this strategy, logic is universal in the sense that the definition of logical validity in set theory applies to all first-order languages, including $\mathscr{L}_{\mathrm{ZF}}$ itself and its first-order expansions.

This is not to say that the model-theoretic definition is an adequate analysis of logical validity for all these languages. Especially when the language of set theory is considered, the absence of interpretations that take quantifiers to range over all sets has worried Kreisel (1967) and many others. In particular, there is no model that can be taken to be the 'intended' interpretation.

The worries can be made more palpable by drawing parallels with restricted versions of the model-theoretic definition. If we concentrate on countable first-order languages only, we could omit all uncountable models as permitted interpretations without actually changing the extension of the consequence relation. This is obvious from the usual Henkin proof of the completeness theorem. However, hardly anyone working in the usual set-theoretic framework would advocate a model-theoretic definition of logical consequence that excludes uncountable models without citing the Löwenheim–Skolem theorem or a version of Kreisel's (1967) squeezing argument (and thus the completeness theorem).

1 The purpose of this paragraph is to sketch the general strategy, not to give a precise account of what Tarski (1936) did. Satisfaction for formulæ of order n with free variables of order $n+1$ is required for the definition of logical validity above. All this was at Tarski's disposal from his earlier work; in (1936) he did not provide formal definitions.

2 (Tarski and Vaught, 1956) is an early paper with the modern definition of a model and the set-theoretic definition of satisfaction in a model.

The restriction to countable (including finite) models becomes a problem when more expressions with a fixed interpretation are added to the language. For instance, once cardinality quantifiers such as ‘there are countably many’, the Härtig quantifier ‘there are as many *A*s as *B*s’, or second-order quantifiers are added, the omission of uncountable models yields an extensionally incorrect definition of the consequence relation. Examples are easy to come by. This is not to say that cardinality quantifiers and so on are actually logical constants; but it should be possible to keep their interpretations fixed without obtaining an inadequate definition of consequence. Definitions of logical consequence that yield inadequate definition of consequence relations when new terms with fixed interpretation are added, are called *unstable* in (Halbach, 2024). Unstable definitions may still yield an extensionally adequate definition when the interpretation of sufficiently many expressions can be varied; but this adequacy then relies on the expressive weakness of the terms whose interpretation is kept fixed.

The usual model-theoretic definition of logical consequence with models of arbitrary cardinality is also unstable. It is stable under the addition of the usual generalized quantifiers, but falters on the addition of the McGee quantifier $\exists^{AI}x\,\phi(x)$ expressing that there are absolutely infinitely (i.e. proper class many) ϕ. As McGee (1992, p. 279) explains, $\exists^{AI}x\,x=x$ is not true in any (set-theoretic) model and thus a contradiction on the model-theoretic definition, although it is true. Thus, the usual model-theoretic definition of logical consequence cannot completely avoid instability, because quantifiers are always interpreted as ranging over the elements of a set.

On the type-theoretic strategy, a higher-order metalanguage is required for the definition of first-order logical truth and consequence. Domains can be and often are omitted from this account. Etchemendy (1999) and others have criticized Tarski for the absence of domains from his (1936) definition of logical consequence. As first-order quantifiers are always interpreted as ranging over all individuals, sentences such as $\exists x\exists y\,x\neq y$ come out as logically valid.[3]

With the rise of first-order set theory as a foundation for mathematics, the model-theoretic definition of logical consequence has become the dominant method. However, Williamson (1999) and others have tried to revive Tarski’s first strategy.[4] They are prepared to bite the bullet by using higher-order logic and accepting the logical truth of $\exists x\exists y\,x\neq y$. The absence of domains means that quantifiers are interpreted as ranging over all (first-order) objects. In particular, the quantifiers of $\mathscr{L}_{\mathrm{ZF}}$ range over all sets. Thus, quantifiers are always interpreted in the same way – as they should if they are logical constants.[5]

I add a remark on my terminology to avoid misunderstandings: The language in which logical consequence is defined needs to be distinguished from the language for which it is defined, although they may coincide in special cases. When I talk about the model- or type-theoretic strategy, I refer to the language in which logical consequence is defined, whatever the language may be for which it is defined. My main focus is on the type-theoretic strategy, applied to the language $\mathscr{L}_{\mathrm{ZF}}$ of first-order set theory. The strategy for defining logical consequence even for type-theoretic languages is nowadays typically model-theoretic. That is, logical consequence for type-theoretic languages is defined in set theory: Interpretations are understood as type structures, which are a special kind of models.

2. Invariance with domains

Obviously both strategies of defining consequence rely on a distinction between the logical and the non-logical vocabulary. On the type-theoretic strategy, only non-logical terms are replaced with variables. On the model-theoretic definition, models assign values only to non-logical terms. For languages such as $\mathscr{L}_{\mathrm{ZF}}$ there is a traditional distinction: In $\mathscr{L}_{\mathrm{ZF}}$ the membership symbol $\in$ is the only non-logical symbol, while connectives and quantifiers are logical. If generalized quantifiers are taken into account, the picture is more blurred. But even if

3 There are some discussions about the exact interpretation of (Tarski 1936). Some authors have denied that Tarski is committed to accepting that $\exists x\exists y\,x\neq y$ is a logical truth. See (Sher, 1996; Gómez-Torrente, 1998, 2009) for a discussion.

4 Logical consequence without domains can also be developed in theories that are not type-theoretic. Friedman (1999) has studied logical consequence without domains in a theory of predication.

5 McGee (2004, p. 374) provides a more thorough discussion. He called quantifiers restricted to domains ‘logical hybrids’ because their interpretation is varied between models, although the interpretation of logical constants should not be varied in the definition of validity. I should define what is meant precisely by saying that the interpretation of a term is kept fixed. In contrast to (Sagi, 2018), I focus here on keeping extensions fixed.

generalized quantifiers are set aside, one would like to have a principled reason for the classification of terms other than tradition. Several ways to make the distinction have been advocated (see, e.g., MacFarlane, 2017, for an overview).

In his (1986) – a talk given in 1966 – Tarski advocated the use of a permutation-invariance criterion for logical notions. This does not directly yield a criterion for classifying terms, that is, syntactic expressions in the language. However, usually it is taken for granted that those terms are logical constants that express logical notions and that it is clear which notion(s) are expressed by a term of the language.[6] I will return to the problem of terms expressing notions later. For now, I proceed as if the transition from logical notions to logical terms were obvious. After the publication of Tarski's paper, invariance criteria for logical notions were further developed in some detail by Sher (1991), McGee (1996), Feferman (1999), Bonnay (2008), and many others (see Bonnay, 2014, for an overview). It is common to all these approaches that they employ domains, as is natural if the distinction is to be applied to the model-theoretic definition of logical consequence.

There is a vast literature on the invariance criteria for logical notions, and there are different attempts to motivate invariance criteria. Tarski (1986) started from the invariance of certain geometric properties under certain transformations and generalized this to invariance under arbitrary permutations. As I mentioned at the beginning, I take the invariance criteria as an attempt to spell out a more traditional characterization of logical constants. Logical 'notions' or operations – the operations expressed by logical terms – should behave indiscriminately on all objects in the same way; they must not behave in a subject-specific way. This requirement is at least partially captured by invariance criteria: A logical operation should behave on objects in the same way as on their proxies (given by some permutation). In the first instance, permutations of a non-empty (set-sized) domain D, that is, an injective mapping of D onto itself, are considered. Then invariance under all permutations of D is defined. Of course, a criterion of logicality is required that is not relative to a domain D: An operation is classified as logical if, and only if, it is logical on all domains.

Usually, relations and operations on D such as identity, relative complementation, intersection, union, and cylindrification are invariant under all permutations over any given domain D. They are thus classified as logical, which is taken as evidence that the terms expressing them – assumed to be the identity, negation, conjunction, disjunction, and existential quantifier symbols – are logical terms relative to every domain. Also cardinality quantifiers like 'There are $\aleph_{17}$ many As' are logical on every domain.

McGee (1996) gave examples that show that a clearly non-logical operation may qualify as logical over certain domains by this criterion. One of them is *wombat disjunction* on p. 575, which is defined as follows for all variable assignments σ:

> σ satisfies $(\phi \mathscr{W} \psi)$ iff either there are wombats in the universe of discourse and σ satisfies either ϕ or ψ or there are no wombats in the universe and σ satisfies both ϕ and ψ.

Wombat disjunction behaves on all domains containing at least one wombat like disjunction and like conjunction on domains without wombats. Since conjunction and disjunction are permutation-invariant on every domain, the operation of wombat disjunction is also permutation-invariant on every domain. If $\mathscr{W}$ expresses the operation of wombat disjunction, it expresses a permutation-invariant operation just like $\wedge$ and $\vee$.

The operation expressed by $\mathscr{W}$ is obviously not logical. The operation of wombat disjunction is sensitive to which objects are in the domain and thus sensitive to the subject matter. Consequently, simple invariance under permutation over arbitrary domains fails to capture the initial informal characterization of logicality as subject independence. Sher (1991) suggested considering not only permutations of domains, but also bijections between domains.[7] McGee (1996) endorsed this generalization to solve the problem of wombat disjunction and other pathological notions. This move is well motivated: A logical operation should work in the same way whether wombats are included in the domain of discourse or not.

[6] At this point many authors remark that (Tarski, 1986) and (Mautner, 1946) have a significant overlap. Tarski was certainly not the first to connect permutation invariance and logicality. Presumably, the are much older predecessors, although they lacked the resources to develop invariance criteria formally. (Mostowski, 1957) also proved to be highly influential.

[7] Mostowski (1957) considered already mappings between domains for generalized quanitfiers, but not for a general characterization of logicality.

However, invariance under all bijections cannot fully solve the problem. Bijections exist only between domains of the same cardinality. A variant of wombat disjunction may not be sensitive to the presence of wombats in the domain, but to the cardinality of the domain. In particular, it may act like disjunction on domains of certain cardinalities, but not others, as (McGee 1996, p. 577) acknowledged: 'The Tarski–Sher thesis does not require that there be any connections among the ways a logical operation acts on domains of different sizes.'

Connections between domains of different cardinalities are established by considering surjective mappings between domains, that is, injectivity is omitted from the admissible mappings between domains. Feferman (1999, 2010) pursued such an approach.[8] If McGee's general framework is retained and bijections are replaced with arbitrary surjective mappings, negation, conjunction, and universal quantification are no longer logical. Feferman obtains the logicality of all the usual logical constants, including negation, conjunction, and universal quantification, of first-order logic without equality by a suitable modification. The main point is the transition from relational (Russellian) type structures to functional type structures. Feferman adds truth values explicitly as elements in the type structure to obtain semantic structures that are in Frege's spirit. But this move trivializes the logicality of the connectives and generates further problems.[9]

The canonical logical notions of first-order logic behave in the same way (in the sense of permutation invariance) on all infinite cardinalities because of the Löwenheim–Skolem theorem. Hence, the problem of different behaviour on domains of different cardinalities does not arise for them. One can now stipulate that a logical notion is 'more logical' if it behaves in the same way on all infinite cardinalities or, at least, on all cardinalities beyond a small cardinal. Sagi (2018) suggested a graded account of logicality in this direction. Kennedy and Väänänen (2021) mapped the resulting classification of logicality. Such an approach means that an additional criterion for logicality beyond permutation invariance is added, namely an independence from the cardinality of the domain. It has led to interesting results; but here I pursue another, more simple-minded strategy for avoiding the problem arising for different behaviour on domains of different size.

3. Invariance over all objects

The complications generated by wombat disjunction and similarly pathological operations have their root in the relativization of the permutation-invariance criterion to domains, which makes it necessary to consider complicated ways for connecting domains.

The relativization of the criterion to domains was motivated by the model-theoretic strategy of defining logical truth and consequence. On the type-theoretic strategy there are no domains, and there is no reason for relativizing invariance criteria to domains. In the remainder of this paper, I argue that, if the criterion of permutation invariance is not relativized to domains, it becomes simple, elegant, and closer to the original aim of capturing independence from the subject matter.

This is by no means a new idea. Already Tarski (1986, p. 149) may have considered both, defining invariance over elements of a specific domain as well as over all objects:

> we would consider the class of all one-one transformations of the space, or universe of discourse, or 'world', onto itself.

Also Williamson (1999) seems to make use of a formulation of the permutation-invariance criterion without domains.

There are obvious obstacles to formulating invariance criteria over all objects. The first difficulty is generated because the extensions of predicates are no longer necessarily sets; they may be proper classes. Even if the extension is a set, to deal with negation, the absolute complement of that set, that is, a proper class, is required. The second obstacle is the need for permutations of the entire universe. Such a permutation cannot be

8 Feferman's characterization of logicality is known as 'homomorphism invariance' approach. There are many subtleties to Feferman's theory, which are not discussed here.

9 Casanovas (2007) provides a more detailed discussion of these two ways of setting up the framework. Here in this paper I focus on operations available in first-order languages and do not discuss (higher-order) type structures.

modelled as a set-sized function. The least tractable objects are the operations themselves; they map proper classes to classes and are thus third-order entities.

I will now formulate the permutation-invariance criterion without domains.[10] Writing out the definitions will help me to assess the resources required – or at least naturally required – for the formulation of the criterion. Operations such as negation or quantification operate on extensions, that is, classes of objects. They will be conceived as functions on extensions and, therefore, as third-order objects. I will define a criterion of permutation invariance only for first-order languages. However, the definition can be extended to higher-order languages; but the definition of permutation invariance for a language of order n necessitates then the ascent to a metalanguage of order $n+2$.

There may be ways to avoid third-order quantifiers, but only with some serious modifications. That third-order logic is needed should not come as a surprise. In the usual setting, type structures over set-sized domains are employed. There, first-order quantifiers are conceived as properties of properties and thus as third-order objects.

A universal definition of logical consequence, in particular, a definition of logical consequence for the language that is being used in that very same language is not within the reach of the proponents of the type-theoretic strategy. Some have left behind all scruples about higher-order quantification and will not mind the use of third-order logic. Those who prefer a lean reading of second-order quantifiers as plural quantification or quantification over predicatively definable classes will find it hard to accept the definitions below, not only because of the use of third-order logic, but also because weak forms of comprehension will make the application of the criterion less plausible.[11]

I choose the language of Zermelo–Fraenkel set theory with urelements as the language for which permutation invariance and finally logical consequence is to be defined because it is a candidate of a language that features quantifiers ranging over all (first-order) objects and has high expressive strength. Of course, this is not what Tarski had in mind in the 1930s; but it is closer to the kind of language formal philosophers of the present day would use as their working language. Together with this language come appropriate axioms. For the sake of definiteness I start with the axioms of Zermelo–Fraenkel set theory with the axiom of choice and an axiom stating that the urelements form a set.[12] I then add full second-order and third-order comprehension. The schemata of first-order Zermelo–Fraenkel are replaced with their higher-order counterparts. Once we accept talk about proper classes, this is a natural way to expand the first-order theory Zermelo–Fraenkel to a higher-order language. This is not a necessary move: For the purposes here, I only need weak assumptions: I define a few notions that require that sequences and functions can be expressed; in the last section I also assume that syntactic facts can be proved in the theory. However, the full development of the theory of logical consequence may be sensitive to whether some strong assumptions are made. For instance, whether the completeness theorem for first-order logic can be proved for languages with quantifiers that are not relativized to domains may depend on the availability of global choice or at least a linear ordering of all first-order objects (Friedman, 1999; Rayo and Williamson, 2003). The existence of a linear ordering can be formulated without any set-theoretic vocabulary; second-order quantification is sufficient.

Urelements do not play a significant role in the formal development. They need to be included because we consider invariance over all objects; and, since there are objects besides the pure sets, I admit urelements. In what follows V is the class of all sets and urelements.

Working informally in the third-order theory, I first define the notion of a variable assignment: A variable assignment is a function from ω; the values of the function can be any (first-order) objects. I write V^ω for the class of all variable assignments. A class of variable assignments is any class $A \subseteq V^\omega$. The variables are

10 McGee (1996) sketched already how to proceed if proper classes are admitted as domains. The approach here is much simpler: There is only one first-order domain, namely the class of all (first-order) objects.

11 In (2020) I consider the formulation of a similar criterion in a pure first-order language. Here I have no ambition to use of a lean metatheory.

12 The referee mentioned in the acknowledgements asked what justifies the assumption that the urelements form a set. The question is tough, because it touches upon question in metaphysics. For instance, propositions may be among the urelements and there may be too many of them to fit into a set. I reject such a conception of propositions. Generally, I think of set theory as the most difficult task that has ever been undertaken for the sake of metaphysics, and nothing can be put on top of the universe of sets, especially not from the realm of more flimsy metaphysics. This is my honest answer; I realize that it is not an argument.

indexed by the elements of ω. I write $v_0, v_1, \ldots$ for these variables. For $a \in V^\omega$ I write $a(k)$ for the value of the variable v_k with index k. I use a and b as variables ranging over variable assignments.

Unlike McGee (1996) for instance, I am only interested in formulæ with finitely many free variables. In particular, I do not admit predicate symbols with infinite arities or infinite conjunctions. They may qualify as logical by some permutation criterion, but the requirement of expressibility in a language with formulæ of finite length overrides the permutation criterion. The use of finite variable assignments leads to notoriously clumsy definitions: For binary connectives, when the variables assignments for the conjoined formulæ are not defined on the same variables, variable assignments have to be merged. It is easier to use variable assignments that are total functions on ω and think of them as finite variable assignments that have been padded out in some way.

The extension of a formula $\phi(v_0, v_2)$, for instance, is understood as the class of all variable assignments that satisfy $\phi(v_0, v_2)$. Since I consider only formulæ with finitely many free variables, all extensions of formulæ are finitary in the following sense:

Definition 1. *A class $A \subseteq V^\omega$ of variable assignments is finitary iff there is a finite set $I \subset \omega$ such that $\forall b \big(\exists a \in A \, \forall i \in I \, b(i) = a(i) \rightarrow b \in A)\big)$.*

The class V^ω itself is a finitary class of variable assignments, as is the empty set $\emptyset$. The class V^ω is the extension of formulæ satisfied by all variable assignments and, in particular, of all true sentences, while $\emptyset$ is the extension of all formulæ not satisfied by any variable assignment and, in particular, of all false sentences.

I could also have employed functions with finite domains as variable assignments. As the extension of a formula $\phi(v_0, v_2)$ with exactly the free variables v_0 and v_2, I could have used the class A_f of all functions from $\{0,2\}$ such that $a(0)$ for v_0 and $a(2)$ for v_2 satisfy $\phi(v_0, v_2)$. Working with such finite variable assignments is more awkward, however. Relative to a set I of (indices of) variables, there is a one-one correspondence between such classes of finite functions with I as domain and the corresponding finitary class of variable assignments. In the example we have $I = \{0,2\}$. The finitary class A corresponding to A_f is obtained by padding out all functions $b \in A_f$ with arbitrary objects:

$$a \in A \text{ iff } \exists b \in A_f \big(b(0) = a(0) \text{ and } b(2) = a(2)\big)$$

That is, we pad out the finite variable assignments in A_f in any possible way. Conversely, restricting all functions in A to the domain $I = \{0,2\}$ yields A_f. This generalizes to other finite index sets I in the obvious way.

I use the term *finitary*, because the restriction to finitary classes of variable assignments corresponds to the restriction that any formula has only finitely many free variables; that is, the arity of any predicate symbol is finite, and there are no infinite conjunctions or predicate symbols with infinite arity.

All non-empty finitary classes of variable assignments are proper classes. In what follows, classes of variable assignments are always assumed to be finitary. However, nothing would go amiss without this restriction. I denote the class of all classes of finitary variable assignments with $\mathscr{F}$. Here it is not necessary to ascend to third-order logic, because having a second-order defining formula for $\mathscr{F}$ will suffice, and thus I can use $\mathscr{F}$ in the same way that set theorists use V or On without committing themselves to proper classes.

The semantic values or extensions of the predicate symbols, the logical connective and quantifier symbols can be understood as operations in the following sense:

Definition 2. *For $n \geq 0$ an n-ary operation O is a function that maps every n-tuple $\langle A_0, \ldots, A_{n-1} \rangle$ of classes of finitary variable assignments to a class of finitary variable assignments A.*

As mentioned above, operations are third-order objects. The operation of negation is a unary operation. It maps a class $A \in \mathscr{F}$ to the complement $V^\omega \setminus A$. The operation of conjunction is binary and maps $\langle A_1, A_2 \rangle$ to $A_1 \cap A_2$. The operation of existential quantification of the k-th variable is unary and maps A to the class

$$\{b : \exists a \in A \, \forall i \neq k \, a(i) = b(i)\}.$$

Note that the values of all these functions are finitary classes of variable assignments, as long as they are applied only to finitary classes of variable assignments.

Strictly speaking, I do not consider relations *simpliciter*, but rather extensions of atomic formulæ. Thus, for instance, there is no single extension for the (class-sized) relation of identity. There are only the classes $Id_{i,j} := \{a : a(i) = a(j)\}$ for all $i \neq j$. Instead of the extension of the formula $v_0 = v_1$, I could consider the binary relation (set of ordered pairs) $Id := \{\langle a(0), a(1)\rangle : a(0) = a(1)\}$, and similarly for other relations. The relation Id is independent of the used variable; that is, in place of 0 and 1, I could have chosen any two variables v_i and v_j with $i \neq j$. Therefore, we can always abstract away from the chosen variables.[13]

The requirement that the indices i and j are distinct cannot be dropped. The formulæ $v_i = v_i$ do not express the identity relation; they do not express any relation, but rather the unary property of being self-identical. In the end, the identity relation and the property of being self-identical qualify as logical; but using the same variable in two distinct occurrences can make a difference to the logicality of the relation or property expressed. For instance, $v_0 \in v_1$ expresses the membership relation (or rather the padded out $A \in \mathscr{F}$) and does not qualify as logical, while $v_0 \in v_0$ expresses the logical empty property $\emptyset$. In definition 5 I formally spell out what it means for a formula to express an operation; but I hope that it is sufficiently clear why $v_0 \in v_1$ and $v_0 \in v_0$ express the respective operations.

The advantage of using classes of variable assignments and thus extensions of formulæ rather than predicate symbols is that it permits a uniform treatment of predicates, connectives, and quantifiers. Classes of variable assignments are just 0-ary operations as in (McGee, 1996). Thus, relations are classes of variable assignments, that is, relations are identified with their extension. In what follows I will be sloppy and talk about *the* operation of identity.

Definition 3. *A permutation of V is an injective mapping of V onto V. The permutation Π' of variable assignments induced by a permutation Π of V is the class-sized function mapping every $a \in V^{\omega}$ to the variable assignment $b \in V^{\omega}$ such that $b(i) = \Pi(a(i))$ for all $i \in \omega$. If Π is a permutation of V and A a class of variable assignments, the permutation $\Pi''(A)$ induced by Π of A is the class $\{\Pi'(a) : a \in A\}$.*

As usual, I conflate permutations of V and the permutations induced by it and write Π where I should write Π' or Π''. It should be clear from the context whether I mean a permutation of V or the induced permutations.

As pointed out above, I am interested in classes of variable assignments in $\mathscr{F}$. Permutations will always stay within $\mathscr{F}$ in the sense that $A \in \mathscr{F}$ implies $\Pi(A) \in \mathscr{F}$.

Permutation invariance can now be defined in the obvious way without any recourse to domains.

Definition 4. *An n-ary operation O is permutation-invariant iff for all permutations Π and all $A_i \in \mathscr{F}$ with $i < n$:*

$$O\langle \Pi(A_1), \ldots, \Pi(A_{n-1})\rangle = \Pi(O\langle A_0, \ldots, A_{n-1}\rangle)$$

The definition yields the expected classification of most operations. All operations of the truth-functional connectives are permutation-invariant, as are the operations of existential and universal quantification and cardinality quantifiers. Also the relation of identity is permutation-invariant. In contrast, the relation $\{a \in V^{\omega} : a(0) \in a(1)\}$ of set-theoretic membership fails to be permutation-invariant, as expected.

Permutation invariance has been defined for operations in general, including those for which our language lacks a primitive symbol. McGee (1992) considered the quantifier $\exists^{AI} v_k$ expressing that there are absolutely infinitely many. This quantifier corresponds to the operation that maps $A \in \mathscr{F}$ to the class B of all variable assignment such that

$$b \in B \text{ iff } (\{a \in A : \forall i \neq k\, a(i) = b(i)\} \text{ is a proper class.})$$

On a straightforward domain-relative formulation of the permutation-invariance criterion, the operation will also qualify as permutation-invariant, but in a trivial way because there is no proper class of variable assignments over a set-sized domain. Without domains the quantifier also qualifies as permutation-invariant. However, it can be shown that the domain-based definition leads to an incorrect classification in contrast to the domain-free

[13] More formally, one could consider the equivalence class of all binary relations $Id_{i,j}$ with $i \neq j$ and then use $Id_{0,1}$ as the representative of the equivalence class. The equivalence class is a second-order definable, and thus there is no real need to ascend to third-order logic. It should be obvious how to generalize this to other relations. I think a referee for helpful notes on this.

definition. The difference can be brought out by considering relativizations of the quantifier. Consider the operation corresponding to $\exists^{AI}v_k\,(On(v_k)\wedge\ldots)$ expressing that there absolutely infinitely many ordinals such that so-and-so. This still qualifies as permutation-invariant on the domain-based account, because it still works like the trivial quantifier $\exists^{AI}v_k\,(v_k\neq v_k\wedge\ldots)$. However, on the domain-free account here, $\exists^{AI}v_k\,(On(v_k)\wedge\ldots)$ fails to be invariant under permutations, as a permutation can map ordinals to non-ordinals. This brings $\exists^{AI}v_k\,(On(v_k)\wedge\ldots)$ in line with other restricted quantifiers such as 'there is at least one cat such that ...' that fail to be logical on both accounts.

Finally, the problem of wombat disjunction is eliminated. The entire problem of connections between domains has vanished because the permutations here are no longer restricted to a domain. Since wombats exist, the operation of wombat disjunction is identical with the operation of disjunction, which is permutation-invariant.

4. Absolutely general quantification

The unrestricted first-order existential and universal quantifiers (or rather the operations expressed by them) are permutation-invariant and thus qualify as logical constants. The domain-based and the domain-free definitions of permutation invariance yield the same results on these quantifiers, and thus is may seem that they treat quantifiers in the same way; but this is misleading.

The application of the traditional domain-based criterion of logical invariance yields the result that *relative to each domain* the universal and existential quantifiers are permutation-invariant. One might object that they are not invariant under expanding or shrinking the domain; for a definitive verdict a clear method for cross-comparisons between different domains would be required. As I have mentioned above, it is far from clear how this can be achieved.

On the domain-free account, quantifiers ranging over a domain D can be mimicked by relativized quantifiers like $\exists v_k\,(v_k\in D\wedge\ldots)$. For fixed k and class D, the relativized quantifier expresses the following operation O_D on all $A\in\mathscr{F}$:

$$O_D(A)=\{b:\exists a\in A\,(\forall i\neq k\,a(i)=b(i)\wedge a(k)\in D)\}$$

The operation O_D is only permutation-invariant if D is empty or the universal class. Thus, on the definition outlined above, the absolutely unrestricted existential and universal quantifiers are permutation-invariant, while the domain-relativized quantifiers are not, because a domain in the sense of model theory is never empty and always a set.[14]

As before, I assume explicitly that whether the symbols $\exists$, $\wedge$, $\neg$, and $=$ are logical constants depends exactly on whether the corresponding operations are. Since then identity qualifies as permutation-invariant, it is hard to avoid an alleged problem in (Tarski, 1936), namely that sentences such as $\exists x\exists y\,x\neq y$ qualify as logical truths.

As the logical validity of these sentences will not be palatable to many, one might try to show that identity is not logical. In fact, Feferman (2010, p. 10) seems to think that the logicality of identity is not decided by invariance approaches:[15]

> Finally, as pointed out to me by Bonnay, it is hard to see how identity could be determined to be logical or not by a set-theoretical invariance criterion of the sort considered here, since either it is presumed in the very notion of invariance itself that is employed – as is the case with invariance under isomorphism or one of the partial isomorphism relations considered in the next section – or it is eliminated from consideration as is the case with invariance under homomorphism.

However, on the account here, distinctness (i.e. non-identity) can be treated as non-logical only by switching from injective to merely surjective mappings. Distinctness is not invariant under such possibly non-injective

[14] Perhaps Frege and others would agree, and only the triumph of model-theoretic semantics since the 1950s makes the claim that domain-restricted quantifiers are not logical sound outlandish. See (McGee, 2004) for a further discussion.

[15] See (van Benthem, 2002, p. 429) for a related circularity objection.

mappings, because objects may be 'merged'. But also conjunction and negation would lose their status as logical constants, because the respective operations are not invariant under all surjective mappings.[16] This is hardly acceptable. Therefore, only invariance under mappings that are surjective *and* injective should be considered; and identity and distinctness are invariant under such bijections.

In response, one could argue that a functional approach as in (Feferman, 1999) could or should be employed; but, as in the case of the domain based approach, that would trivialize the logicality of the connectives, because their values would be truth values that are unaffected by permutations.[17]

Therefore it is hard to see why identity and non-identity should not be logical constants. $\exists x \exists y\, x \neq y$ is true and does not contain any non-logical constant that can be re-interpreted. Hence, it is logically valid. If that is rejected, another criterion not solely based on invariance is needed.

5. Expressing operations

Even if a criterion for distinguishing between logical and non-logical operations is available, we still need to distinguish between logical and non-logical terms or expressions in the language. Only if we have a criterion for deciding which operations are expressed by terms in the language, will a criterion for operations be of use in the definition of logical consequence. This is because for the definition of logical truth and consequence, we must know which terms can be re-interpreted and which ones have an interpretation that is kept fixed. Only atomic, not complex, expressions need to be classified for the purpose of defining logical truth and consequence. Therefore, by 'term' I always mean 'atomic term'.

The distinction between atomic and complex expressions is not trivial. Predicate symbols should qualify as atomic, as should connective and quantifier symbols. By 'atomic' I do not mean 'syntactically atomic'. Arguably, in standard infix notation the brackets in $(\phi \wedge \psi)$ belong to the term of conjunction. One might like to say that an expression is atomic if it has a meaning of its own; but that does not clarify much. At any rate, the distinction should be sufficiently clear for the standard first-order languages.

Criteria of invariance such as the one above only yield a criterion of logicality for operations if it is stipulated that an operation is logical iff it is invariant under the relevant mappings. To apply the criterion to terms one would like to say that a term is logical iff the operation expressed by the term is permutation-invariant. The guiding idea behind much of the literature on permutation invariance and in the section above has been the following definition:

Definition of logical constants. *A term is logical iff is expresses a permutation-invariant operation.*

Of course, it needs to be made precise what it means for a term to express an operation. Still working in third-order set theory, it can be made precise for use with my invariance criterion by defining a formula $\mathsf{Sat}(x,y)$ in second-order set theory expressing that the first-order formula x is satisfied by the variable assignment y.

Definition 5. *Assume that $\circ$ is a predicate symbol or an n-ary connective or quantifier and define $|\phi| := \{a \in V^{\omega} : \mathsf{Sat}(\ulcorner \phi \urcorner, a)\}$. Then $\circ$ expresses an operation O that maps $\langle |\phi_1|, \ldots, |\phi_n| \rangle$ to $|\circ(\phi_1, \ldots, \phi_n)|$ for all first-order formulæ of the chosen language.*

This contains the special case of atomic formulæ: An atomic formula $Rv_{i_1} \ldots v_{i_n}$ expresses the 0-ary operation $\{a \in V^{\omega} : \mathsf{Sat}(\ulcorner Rv_{i_1} \ldots v_{i_n} \urcorner, a)\}$. Consequently, extensionality is built into this account: Whether a formula expresses a permutation-invariant relation depends only on the objects to which it applies, not on any other features of the formula.

As I have mentioned above, atomic formulæ $Rv_{i_1} \ldots v_{i_n}$ and $Rv_{j_1} \ldots v_{j_n}$ express different operations if $i_k \neq j_k$ for some $k \leq n$ and R does not happen to apply to all objects in the k-th argument place. However, as explained above, we can abstract away from the specific variables. The operations expressed by $Rv_{i_1} \ldots v_{i_n}$ and $Rv_{j_1} \ldots v_{j_n}$ will both be permutation-invariant (or not) as long as for every $k, l \leq n$, $i_k = i_l$ iff $j_k = j_l$. The

16 See again (Casanovas, 2007).
17 See again (van Benthem, 2002, p. 429).

last condition means that in $Rv_{i_1}\dots v_{i_n}$ the same argument places must occupied by identical variables as in $Rv_{j_1}\dots v_{j_n}$. The example of membership above provides a counterexample to the claim without this restriction: $v_0 \in v_1$ expresses an operation that fails to be permutation-invariant, while the operation expressed by $v_0 \in v_0$ is permutation-invariant. In the first formula the two variables are distinct, while they coincide in the second. This is what one would intuitively expect: $v_0 \in v_1$ expresses membership, $v_0 \in v_0$ the empty relation, at least is standard set theory with foundation.

The definition of expressing an operation is relative to the language of the formulæ $\phi_1,\dots,\phi_n$. There may be sets of finitary variable assignments that cannot be defined in the given language. That is, there may be (and there will be) sets of variable assignments that are not identical to any $|\phi|$. The definition thus does not impose any restrictions on how the operation has to behave on undefinable sets of finitary variable assignments. Thus $\circ$ will express more than one operation.

It is not sufficient to define $|\phi| := \{a \in V^\omega : \phi(a(k_1),\dots,a(k_n))\}$ for ϕ containing exactly $v_{k_1},\dots,v_{k_n}$ free. This would yield only a schema. The problem is that in the definition I quantify over formulæ; but in $\{a \in V^\omega : \phi(a(k_1),\dots,a(k_n))\}$ the formula is used, not mentioned, and can thus not be quantified.

Definition 5 is not available on the standard model-theoretic account with domains, because Sat cannot be defined. Only satisfaction relative to a set-sized model is definable. Moreover, $|\phi|$ is always a proper class unless it is empty. In informal discussions philosophers do make use of a satisfaction predicate that presumably belongs to a not further specified metatheory. It should be clear that this metatheory cannot be Zermelo–Fraenkel set theory, but some properly stronger theory.

For the usual predicate, connective, and quantifier symbols this definition yields the expected result: $v_0 = v_1$ expresses the relation, that is, the 0-place operation $Id_{0,1}$ of identity above, while the symbol $\neg$, for instance, expresses the operation that maps A to its complement $V^\omega \setminus A$.[18]

After having given the definition of logical constants a precise form, or at least a form that can easily be made precise, I conclude with a few remarks about its adequacy and how it can be tested by introducing new terms into the language. There are counterexamples against the adequacy of the right-to-left direction of the definition, that is, there are non-logical terms expressing a permutation-invariant operator. Various such examples have been discussed in the literature in the context of domain-based invariance criteria. Wombat disjunction is one of them; but, of course, that is no longer a problem once we move to the domains-free account.

McGee (1996, p. 569) gave another example, unicorn negation $\mathscr{U}$, which expresses the same operation as negation:

$$\mathscr{U}\phi := (\neg\phi \wedge \text{there are no unicorns}),$$

Since there are no unicorns, the two connectives $\neg$ and $\mathscr{U}$ express the same operation. Assuming that a term is logical if, and only if, it expresses a permutation-invariant operation, as in my definition above, both, $\neg$ and $\mathscr{U}$, are therefore logical constants. There is no way that a criterion based on invariance can distinguish between the two, although one would expect $\neg$ to be a logical constant while $\mathscr{U}$ is not.

One might try to argue that unicorn negation poses the same problem as wombat conjunction by insisting that there are domains with and without unicorns; the only difference would be that unicorns are mere *possibilia*. However, if we follow Kripke (1980), there are no unicorns by necessity, and it will not help to consider other metaphysically possible worlds than our own, as they also lack unicorns. If we do not follow Kripke in this and believe that there are possible unicorns, we can replace 'unicorns' with 'heptahedrons' in the sense of 'regular polyhedron of seven faces', which do not exist by mathematical necessity (Gómez–Torrente, 2002, p. 18). The point is that the second conjunct is not true in any domain.

Examples of potentially logical terms such as unicorn negation introduced via definitions have their problems. Sagi (2015) turned the attention to the way unicorn negation is introduced. She points out that the

[18] To show that, e.g., negation expresses complementation relative to $\mathscr{F}$, one will use $\forall\phi\,\forall a\,(\mathsf{Sat}(\ulcorner\neg\phi\urcorner, a) \leftrightarrow \neg\mathsf{Sat}(\ulcorner\phi\urcorner, a))$. Bernays–Gödel set theory can define a truth predicate that yields the T-sentences for set-theoretic sentences; but such a truth predicate cannot provably commute with connective and quantifiers (see Halbach, 2014, p. 19f). This is the main reason to employ full second-order comprehension as in Morse–Kelley, and not only a conservative theory such as Bernays–Gödel set theory.

example is unconvincing if the symbol ':=' in the definition of unicorn negation is understood as expressing that $\mathscr{U}\phi$ is merely an abbreviation of the right-hand side. If it were a mere abbreviation, I could tweak my definitions above. Probably the most obvious way would be to change the definition of (atomic) term; and an expression such as $\mathscr{U}$ should be treated as a complex expression and thus not be a candidate for being a logical constant. Alternatively, one could try to declare such abbreviations as $\mathscr{U}$ logical if, and only if, all terms in the definiens are logical. Whether unicorn negation is logical then depends on whether 'is a unicorn' is logical or not. As the predicate 'is a unicorn' does not apply to anything, it qualifies as logical under the permutation criterion.[19]

At any rate, I doubt that we should think of unicorn negation or wombat conjunction as having been introduced as mere abbreviations; at least we do not have to think of them as mere abbreviations. Introducing terms like unicorn negation or the McGee quantifier $\exists^{AI}x$ by definition is a somewhat confusing affair. We can easily introduce a new logical term if the definition contains only logical terms: For instance, we can introduce a new atomic symbol for distinctness from negation and identity, or for ternary conjunction from binary conjunction. If a new term is defined from terms that are logical by the above criterion, one can prove that the new term is also logical. However, these are less interesting examples. In the interesting cases a new term is introduced by a definiens containing non-logical terms, and it is not possible to give a definition with merely logical terms. For instance, the quantifier 'there are at least $\aleph_5$-many *A*s' can be introduced only by a complex set-theoretic formula, at least as long as we are in the usual expansions of the language of set theory. The complex formula, and presumably even an atomic expression merely abbreviating it, are not logical constants, because membership is not logical by any invariance criterion. However, one can understand the definition as a way of picking out a certain quantifier whose meaning then no longer depends on the chosen defining formula.

Defining new terms that may qualify as logical constants resembles in some aspects the introduction of a designator using Kaplan's (1978) *dthat*, which generates a directly referring term from a description. For the introduction of new quantifiers we would need an operator that generates a new primitive quantifier from a given complex formula.[20]

If we were able to introduce new quantifiers by such definitions, there would be two main kinds of formulæ in the language for stating that there are at least $\aleph_5$-many self-identical objects: One could use either a set-theoretic formula or the newly introduced atomic quantifier. In the former case the resulting formula contains non-logical vocabulary, in particular the membership symbol; in the latter it contains only logical vocabulary, because the quantifier expresses a permutation-invariant operation. Under the type-theoretic theory of logical consequence, the latter will be logically valid while the former is not. This does not seem unreasonable. Perhaps even the definition of unicorn negation can be understood this way, merely as a way to fix a logical operation (although, clearly, this is not how McGee intends it to be understood).

If I were to introduce the McGee quantifier $\exists^{AI}x$ in a proper way, I would resort to such a definition that is not merely an abbreviation. $\exists^{AI}x\,\phi(x)$ would be defined as $\neg\exists y \forall x\,(x \in y \leftrightarrow \phi(x))$ expressing that there is no set of all ϕs; but $\exists^{AI}x\,\phi(x)$ would not be a mere abbreviation of the formula; the set-theoretic formula $\neg\exists y \forall x\,(x \in y \leftrightarrow \phi(x))$ is only used to pick out the operation of the quantifier. Thus, $\exists^{AI}x\,x = x$ is at least a candidate for being a logical truth, while $\neg\exists y \forall x\,(x \in y \leftrightarrow x = x)$ is not and only a set-theoretic truth.[21]

I do not think that artificial examples with defined terms such as unicorn negation are required as counterexamples to the adequacy of logical constants given above. The predicate 'is a unicorn' is already an example. It expresses a permutation-invariant operation and is thus logical, according to the criterion. There is no reason to

[19] Sagi (2015, p. 167) seems to suggest that McGee does not give us any reason to think that 'is a unicorn' is logical. This is correct. I suppose that McGee thought that its logicality is thus obvious, if an invariance criterion is applied.

[20] A referee suggested the analogy to *dthat*. I had used the definition of Neptune as the planet that is causing the perturbations in the orbit of Uranus (Kripke 1972). Here 'Neptune' is not a mere abbreviation of the definite description, but directly referring to the planet. All these analogies have their limits. In particular, the notion of a directly referring term can be made precise by using metaphysically possibility, which does not appear to be of much use here.

[21] (Sagi 2015, p. 160) mentions that we could think of $\mathscr{U}$ as having been introduced by a rule in the metalanguage. At least I do not think it should be introduced by fixing how $\mathscr{U}$ should be interpreted in a given model. The McGee quantifier $\exists^{AI}$ could definitely not be introduced in this way, because every sentence $\exists^{AI}x\,\phi(x)$ is false in all models. What is missing from Sagi's account is a third way of introducing new vocabulary: A definition in the object language need not be just an abbreviation. This is well-known from examples such as '*dthat*', Neptune, and standard metre bar examples examples.

assume that 'is a unicorn' need or can be introduced by a definition of any kind. Unicorn negation is merely an attempt to provide an example that is a connective; but predicates not satisfied by anything are already a problem for the usual invariance accounts of logicality.

There are duller examples: The sentence parameter or 0-place predicate P could happen to be false; it could stand for something like 'it is raining'. No criterion of permutation invariance can tell us that the 0-place connective $\perp$ for *falsum* is a logical constant and P is not, because the semantic status of neither depends on any object.

Therefore, as it stands, I do not endorse the definition of logical constants above. It is an improvement on the domain-based criterion; but it still yields only a necessary criterion for the logicality of an expression in the language.[22]

Acknowledgements

I thank Kentaro Fujimoto, Beau Mount, Lavinia Picollo, Philip Welch, and Tim Williamson for illuminating discussions of his topics as well as two referees of this journal and a referee of another journal for their insightful comments. I am also very grateful to the Leverhulme Trust for supporting my research with a Philip Leverhulme Research Fellowship.

References

Bonnay, D. (2008). Logicality and invariance. *Bulletin of Symbolic Logic*, 14, 29–68.
Bonnay, D. (2014). Logical constants, or how to use invariance in order to complete the explication of logical consequence. *Philosophy Compass*, 9, 54–65.
Casanovas, E. (2007). Logical operations and invariance. *Journal of Philosophical Logic*, 36, 33–60.
Etchemendy, J. W. (1999). *The Concept of Logical Consequence* (2nd ed.). CSLI, Stanford.
Feferman, S. (1999). Logic, logics, and logicism. *Notre Dame Journal of Formal Logic*, 40, 31–54.
Feferman, S. (2010). Set-theoretical invariance criteria for logicality. *Notre Dame Journal of Formal Logic*, 51, 3–20.
Friedman, H. (1999). A complete theory of everything: Satisfiability in the universal domain. Retrieved from https://u.osu.edu/friedman.8/files/2014/01/ACompThyEver101099-1mkg42b.pdf
Gómez-Torrente, M. (1998). On a fallacy attributed to Tarski. *History and Philosophy of Logic*, 19, 227–234.
Gómez-Torrente, M. (2002). The problem of logical constants. *Bulletin of Symbolic Logic*, 8, 1–37.
Gómez-Torrente, M. (2009). Rereading Tarski on logical consequence. *Review of Symbolic Logic*, 2, 249–297.
Halbach, V. (2011). *Axiomatic Theories of Truth* (1st ed.). Cambridge University Press, Cambridge.
Halbach, V. (2014). *Axiomatic Theories of Truth* (revised ed.). Cambridge University Press, Cambridge.
Halbach, V. (2020). The substitutional analysis of logical consequence. *Noûs*, 54, 431–450.
Halbach, V. (2024). *The Definition of Logical Consequence*. Book manuscript.
Kaplan, D. (1978). Dthat. In P. Cole (Ed.), *Pragmatics, Syntax and Semantics* (pp. 221–243). Academic Press.
Kennedy, J, & Väänänen, J. (2021). Logicality and model classes. *Bulletin of Symbolic Logic*, 27, 385–414.
Kreisel, G. (1967). Informal rigour and completeness proofs. In I. Lakatos (Ed.), *The Philosophy of Mathematics* (pp. 138–171). North Holland.
Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *The Semantics of Natural Languages* (pp. 253–355). Reidel.
Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
MacFarlane, J. (2017). Logical constants. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.). Metaphysics Research Lab, Stanford University.
Mautner, F. I. (1946). An extension of Klein's Erlanger program: Logic as invariant-theory. *American Journal of Mathematics*, 68, 345–384.
McGee, V. (1992). Two problems with Tarski's theory of consequence. *Proceedings of the Aristotelian Society*, 92, 273–292.
McGee, V. (1996). Logical operations. *Journal of Philosophical Logic*, 25, 567–580.
McGee, V. (2004). Tarski's staggering existential assumptions. *Synthese*, 142, 371–387.
Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, 44, 12–36.
Rayo, A. & Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In *Liars and Heaps: New Essays on Paradox* (pp. 331–356). Oxford University Press.

[22] See (McGee, 1996, p. 570) for a similar assessment of permutation invariance criteria for classifying terms in the language as logical.

Sagi, G. (2015). The modal and epistemic arguments against the invariance criterion for logical terms. *Journal of Philosophy*, 112, 159–167.

Sagi, G. (2018). Logicality and meaning. *The Review of Symbolic Logic*, 11, 133–159.

Sher, G. (1991). *The Bounds of Logic: A Generalized Viewpoint*. MIT Press.

Sher, G. (1996). Did Tarski commit 'Tarski's Fallacy'? *Journal of Symbolic Logic*, 61, 653–686.

Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica Commentarii Societatis Philosophicae Polonorum*, 1, 261–405. Translated as The Concept of Truth in Formalized Languages in Tarski (1956), 152–278.

Tarski, A. (1936). Über den Begriff der logischen Folgerung. *Actes du congrès international de philosophie scientifique*, 7, 1–11. Translated as On the Concept of Logical Consequence in Tarski (1956), 409–420.

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

Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, 7, 143–154. Edited by John Corcoran.

Tarski, A, & Vaught, R. (1956). Arithmetical extensions of relational systems. *Compositio Mathematica*, 13, 81–102.

van Benthem, J. (2002). Logical constants, the variable fortunes of an elusive notion. In W. Sieg, R. Sommer, & C. Talcott (Eds.), *Reflections on the Foundations of Mathematics: Essays in Honor of Solomon Feferman* (Vol. 15 of Lecture Notes in Logic, pp. 420–440). Cambridge University Press.