UNRESTRICTED QUANTIFICATION AND LOGICAL CONSTANTS

Volker Halbach*

13th June 2022

TWO STRATEGIES FOR DEFINING LOGICAL CONSEQUENCE

Two strategies for defining logical truth and consequence have their origin in Tarski's work. The first 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* Tarski (1935). On this strategy, a formula A is classified as logically true if, and only if the result of replacing all non-logical terms with appropriate variables 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 variable assignments. The techniques developed by Tarski in *The concept of truth* yield a definition of satisfaction for formula of order n in a language of order $n+1$ relative to a variable assignment that can be used for this purpose.[1] No mention is made of domains of quantification on the type-theoretic strategy.

The second strategy devised by Tarski is the 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

*I thank Kentaro Fujimoto, Beau Mount, Lavinia Picollo, and Tim Williamson for discussions of his topic. I am also very grateful to the Leverhulme Trust for supporting my research with a Philip Leverhulme Prize.

[1]The purpose of this paragraph is to sketch the general strategy, not to give a precise account of what Tarski (1936) did. Tarski, for instance defined logical consequence, not logical validity of a single sentence. Moreover, 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; but in (1936) he did not provide formal definitions.

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 defined. 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 strategies. Here in this paper I focus on definitions of logical consequence for first-order languages with a special interest in the language $\mathcal{L}_{\mathrm{ZF}}$ of set theory and its expansions, because they are the languages that are candidates as regimented versions of the language many logicians, including myself, use for theorizing. The model-theoretic definition has the advantage that it does not require an ascent to a higher-order language. Thus logic is universal on this strategy in the sense that the definition of logical validity in set theory applies to all first-order languages, including $\mathcal{L}_{\mathrm{ZF}}$ itself and its first-order expansions.

On the type-theoretic strategy, a second-order metalanguage is required for the definition of first-order logical truth and consequence. The absence of domains has claimed to be problem. Etchemendy (1999) and others criticized Tarski for omitting domains on the type-theoretic strategy, which makes sentences such as $\exists x\, \exists y\, x \neq y$ logically valid, as quantifiers are always interpreted as ranging over all objects.[3] With the rise of first-order set theory as a foundations 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. They are prepared to bite the bullet by using higher-order logic and accepting the logicality of $\exists x\, \exists y\, x \neq y$. The absence of domains means that quantifiers can be interpreted as ranging over all (first-order) objects. In particular, the quantifiers of $\mathcal{L}_{\mathrm{ZF}}$ can range over all sets. Thus, quantifiers are interpreted always in the same way – as they should if they are logical constants.[4] On the model-theoretic account there cannot be an intended interpretation, as quantifiers can only range over the elements of a set.

[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.

[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).

[4] 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.

Both strategies 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 $\mathcal{L}_{\text{ZF}}$ there is a traditional distinction: In $\mathcal{L}_{\text{ZF}}$ the membership symbol $\in$ is the only non-logical symbol, while connectives and quantifiers are logical. Once generalized quantifiers are taken into account, the picture is more blurred. But even if generalized quantifiers are set aside, one would like to have a 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 (1986) – a talk given in 1966 – Tarski advocated the use of a permutation invariance criterion for logical terms, that is, logical constants. Subsequently, invariance criteria for logicality were 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.

Logical operations – the operations expressed by logical terms – should behave 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 in the same way on objects as it does 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, is considered. Then invariance under all permutations of D is defined. Usually, relations and operations on D such as identity, relative complementation, intersection, union, and cylindrification are 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 *A*s' are logical on every domain.

Of course, a criterion of logicality is required that is not relative to a domain, because for the definition of logical consequence an absolute distinction is needed. It is customary to define the absolute notion of logicality in the obvious way: An operation is classified as logical if, and only if it is logical on all domains.

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 $(\phi \mathcal{W} \psi)$ 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 every domain, also the operation of wombat disjunction is permutation-invariant on every domain. If $\mathcal{W}$ expresses the operation of wombat disjunction, it expresses a permutation invariant operation just like $\wedge$ and $\vee$.

However, $\mathcal{W}$ is not logical. The operation of wombat disjunction is sensitive to which objects are in the domain. It is 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 to consider not only permutations of domains, but also bijections between domains.[5] McGee (1996) endorsed this generalization to solve the problem of wombat disjunction and other pathological notions. This move is well motivated: A logical operations should work in the same way whether wombats are included in the domain of discourse or not.

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 domains 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.[6] 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 at least and generates further problems.[7]

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

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

[7] 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

The complications generated by wombat disjunction and similarly pathological operations are caused by relativizing 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.

Williamson's account in (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 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.[8] Writing out the definitions will help me to assess the resources required or at least naturally required to formulate 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 $n + 2$.

(higher-order) type structures.

[8] 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.

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 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.[9]

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 axioms of choice and an axiom stating that the *urelements* form a set. I then add full second-order and third-order comprehension. The schemata of first-order Zermelo–Fraenkel are replaced with their higher-order counterparts. 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).

Urelements do not play a significant rôle in the formal development. However, they need to be included because we consider invariance over all objects. Since there are objects besides the pure sets, I admit *urelements*. I assume that there is a set of all *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

[9] In XXX I consider the formulation of a similar criterion in a pure first-order language. Here my ambition is not the use of a lean metatheory.

can be any 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 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.

I think of the extension of a formula $\phi(v_0, v_2)$, for instance, as the class of all variable assignments hat satisfy $\phi(v_0, v_2)$. A finite variable assignment for the formula would be a function from $\{0, 2\}$; and its extension would be the class A of all such functions such that $\langle a(0), a(2)\rangle$ satisfies $\phi(v_0, v_2)$. The result of padding out A is the class of all functions $b \in V^\omega$ such that for some $a \in A$, $b(0) = a(0)$ and $b(2) = a(2)$. I focus on classes of such variable assignments:

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

I use the term *finitary*, because the restriction to finitary classes of variable assignments corresponds ot 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 conjunction or the like.

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 $\mathcal{F}$. Here it is not necessary to ascend to third-order logic, because having a predicate for $\mathcal{F}$ will suffice. I can use $\mathcal{F}$ is the same way that set theorists use V or On without committing themselves to proper classes.

The semantic values or extensions of predicate symbols, 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 $A \in \mathcal{F}$ 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.

Relations, the semantic values of predicate symbols, are identified with 0-ary operations as in (McGee 1996). Thus, relations are classes of variable assignments, that is, relations are identified with their extension. There is, however, an annoying dependence on indices. Consider, for instance, the relation of identity:

$$Id_{0,1} := \{a : a(0) = a(1)\}$$
$$Id_{2,7} := \{a : a(2) = a(7)\}$$

The operation $Id_{0,1}$ is the extension of the formula $v_0 = v_1$, while $Id_{2,7}$ is the extension of $Id_{2,7}$. Both are operations of identity. However, the difference is merely a renaming of variables. 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 variable assignments in $\mathcal{F}$. Permutations will always stay within $\mathcal{F}$ in the sense that $A \in \mathcal{F}$ implies $\Pi(A) \in \mathcal{F}$.

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

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

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

Now the definition yields the expected classification of most operations. All operations of the truth-functional connectives are permutation-invariant, as are the operation 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 $A \in \mathcal{T}$ 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 sized classes 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 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 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 $\exists v_k\ (v_k \neq v_k \wedge \ldots)$. 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, wombat disjunction behaves exactly like normal disjunction, which is permutation-invariant. This may be seen as problematic, and I will add some remarks in the final section.

ABSOLUTELY GENERAL QUANTIFICATION

The unrestricted first-order existential and universal quantifiers are permutation-invariant and thus qualify as logical constants. It may thus seem that the domain-based and the domain-free definitions of permutation invariance yield the same results on these quantifiers; 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 $\exists x\ (x \in D \wedge \ldots)$. These relativized quantifiers are 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.[10]

For the rest of this section I assume explicitly that whether the symbols $\exists$, $\wedge$, $\neg$, and $=$ depends exactly on whether the corresponding operations are. Since then identity qualifies as permutation-invariant, it is hard to avoid 'Tarski's fallacy', that is, the logical validity of sentences such as $\exists x \, \exists y \, x \neq y$.

As the logical validity of these sentences will not be palatable to many, one might try to disqualify identity from being a permutation-invariant and thus, presumably, logical. In fact, (Feferman 2010, p. 10) seems to think that the logicality of identity is not decided by invariance approaches:

> 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, non-identity cannot be treated as non-logical only by switching from injective to merely surjective mappings. Non-identity is not invariant under such possibly non-injective 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. This is hardly acceptable. Therefore, only invariance under injective mappings should be considered; and identity and non-identity are invariant under such mappings.

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 cased approach, that would trivialize the logicality of the connectives, because their values would be truth values that are unaffected by permutations.

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.

EXPRESSING OPERATIONS

Even if a criterion for distinguishing between logical and non-logical operations is available, we still need ot distinguish between logical and non-logical terms or expres-

[10] 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.

sions in the language. A criterion for deciding whether terms such as the symbols for negation, identity, and set membership are logical or not. For the definition of logical truth and consequence, we must know which terms can be re-interpreted and whose interpretation is kept fixed. Only atomic, not complex expressions need not to be classified for the purpose of defining logical truth and consequence. Therefore, by 'term' I always understand '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. I would 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 yield only 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 still 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 the 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 the 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 $\{a \in V^{\omega} : \mathsf{Sat}(\ulcorner \phi \urcorner, a)\}$. Consequently, extensionality is built into this account: Whether a formula expresses an permutation-invariant relation depends only the objects to which it applies, not on any other features of the formula.

The definition is relative to a language admits that a term expresses more than one operation. There may be finitary variable assignments that cannot be defined in he given language. The definition thus does not impose any restrictions on how the operation has to behave on undefinable finitary variable assignments.

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.

The definition 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; but it should be clear that this metatheory cannot be Zermelo–Fraenkel set theory; but some theory properly stronger.

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 $V^\omega \setminus A$.[11]

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 the its adequacy. There are counterexamples against the right-to-left direction of the definition; and there may be non-logical terms expressing a permutation-invariant operator.

As a first simple example consider the propositional constant $\bot$ for *falsum* and the sentence parameter or 0-place predicate P, and assume that P happens to be false. No criterion of permutation invariance can tell us that $\bot$ is a logical constant and P is not, because the semantic status of neither depends on any object.

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

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

Therefore, assuming that a term if logical iff it expresses a permutation-invariant operation, both $\neg$ and $\mathcal{U}$ are logical constants.

Examples of this kind have their problems. Sagi (2015) has rightly turned the attention to the way unicorn negation is introduced. Sagi points out that the example is unconvincing if ':=' in the definition of unicorn negation is understood as a mere abbreviation or definition in the object language. If this is correct, I need to tweak my definitions above. Probably the most obvious way would be to change the definition of

[11]To show that, e.g., negation expresses complementation relative to $\mathcal{F}$, 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 a truth predicate 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.

(atomic) term; and an expression such as $\mathcal{U}$ should not be treated as a term. However, I do not know how to give a precise definition of what should count as an atomic expression that would give the expected result.

Introducing logical constants by definition is a somewhat confusing affair. Of course, we can easily introduce a new logical term if the definition contains only logical terms: We can introduce a new atomic symbol for distinctness from negation and identity, or for ternary conjunction from binary conjunction. However, in the interesting cases this is not possible. 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 even an atomic expression merely abbreviating it, are not logical constants. However, one can understand the definition merely as a way of picking out a certain quantifier whose meaning then no longer depends on the chosen defining formula. Defining new logical constants in this way resembles in some aspects the introduction of rigid designators by a definite description, for instance 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.

If we were able to introduce a logical constant in this way, there would be two ways in the language to state that there are at least $\aleph_5$-many self-identical objects. One could use either the 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. Under a theory of logical consequence as in (Williamson 1999), the latter will be logically true while the former is not. We could go further and replace the long set-theoretic formula with a single symbol that is understood as a mere abbreviation. The claim that there are $\aleph_5$-many self-identical objects will still not be logical, if formulated with this single symbol. At any rate, the criterion for logicality, as stated above, will struggle to distinguish between the logical constant and the symbol abbreviation the long formula.

It is obvious that a simple criterion like the one above cannot be sensitive to how a quantifier has been introduced. Therefore, as it stands, I do not endorse the definition of logical constants above. At best, it yields a necessary criterion for the logicality of an expression in the language.

References

Bonnay, Denis (2008), 'Logicality and invariance', *Bulletin of Symbolic Logic* 14, 29–68.

Bonnay, Denis (2014), 'Logical constants, or how to use invariance in order to complete the explication of logical consequence', *Philosophy Compass* 9, 54–65.

Casanovas, Enrique (2007), 'Logical operations and invariance', *Journal of Philosophical Logic* 36, 33–60.

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

Feferman, Solomon (1999), 'Logic, logics, and logicism', *Notre Dame Journal of Formal Logic* 40, 31–54.

Feferman, Solomon (2010), 'Set-theoretical invariance criteria for logicality', *Notre Dame Journal of Formal Logic* 51, 3–20.

Friedman, Harvey (1999), 'A complete theory of everything: satisfiability in the universal domain'.
URL: *https://u.osu.edu/friedman.8/files/2014/01/ACompThyEver101099-1mkg42b.pdf*

Gómez-Torrente, Mario (1998), 'On a Fallacy Attributed to Tarski', *History and Philosophy of Logic* 19, 227–234.

Gómez-Torrente, Mario (2009), 'Rereading Tarski on Logical Consequence', *Review of Symbolic Logic* 2, 249–297.

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

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

MacFarlane, John (2017), Logical constants, *in* E. N.Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Winter 2017 edn, Metaphysics Research Lab, Stanford University.

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

McGee, Vann (1996), 'Logical operations', *Journal of Philosophical Logic* 25, 567–580.

McGee, Vann (2004), 'Tarski's staggering existential assumptions', *Synthese* 142, 371–387.

Mostowski, Andrzej (1957), 'On a generalization of quantifiers', *Fundamenta Mathematicae* 44, 12–36.

Rayo, Agustín and Timothy Williamson (2003), A completeness theorem for unrestricted first-order languages, *in* 'Liars and Heaps: New Essays on Paradox', Oxford University Press, pp. 331–356.

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

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

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

Tarski, Alfred (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); page references are given for the translation.

Tarski, Alfred (1936), '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 (1956), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Clarendon Press, Oxford.

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

Tarski, Alfred (1986), *Collected papers*, Vol. 1, Birkhäuser, Basel.

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

Williamson, Timothy (1999), 'Existence and contingency', *Aristotelian Society* sup. vol. 73, 227–234. Reprinted with printer's errors corrected in Proceedings of the Aristotelian Society 100 (2000): 321–343 (117–139 in unbound version).