The Bulletin of Symbolic Logic
Volume 28, Number 1, March 2022

INVARIANCE CRITERIA AS META-CONSTRAINTS

GIL SAGI

Abstract. Invariance criteria are widely accepted as a means to demarcate the logical vocabulary of a language. In previous work, I proposed a framework of "semantic constraints" for model-theoretic consequence which does not rely on a strict distinction between logical and nonlogical terms, but rather on a range of constraints on models restricting the interpretations of terms in the language in different ways. In this paper I show how invariance criteria can be generalized so as to apply to semantic constraints on models. Some obviously unpalatable semantic constraints turn out to be invariant under isomorphisms. I shall connect the discussion to known counter-examples to invariance criteria for logical terms, and so the generalization will also shed light on the current existing debate on logicality. I analyse the failure of invariance to fulfil its role as a criterion for logicality, and argue that invariance conditions should best be thought of as merely methodological meta-constraints restricting the ways the model-theoretic apparatus should be used.

§1. Introduction. Tarski's model-theoretic definition of logical consequence relies on a distinction between the terms in a language that have a fixed interpretation, and those whose interpretation varies across models. Logical terms are those whose meanings get fixed across models. Recent work on logicality is thus aimed at completing the definition of logical consequence, telling us which terms ought to be fixed. Invariance criteria, in particular, have proven to be especially fruitful and interesting, integrating philosophical and mathematical studies of logicality.

In this paper I reconsider the role of invariance criteria. I take here a wider perspective, from which logical consequence is no longer defined through a division of terms into logical and nonlogical, and invariance criteria will be generalized accordingly. It will emerge that when it comes to completing the definition of logical consequence, invariance criteria are unsatisfactory, and the generalization will direct us to the issues that need to be resolved for an adequate solution to the problem of logicality.

In recent work, I proposed a framework for model-theoretic consequence which does not rely on a strict distinction between logical and nonlogical terms, but rather on a range of *semantic constraints* on models restricting the interpretations of terms in the language in different ways—not by merely either fixing their interpretation or leaving it variable. An example

Received September 11, 2019.
2020 *Mathematics Subject Classification*. 03A05, 03C99.
Key words and phrases. logical consequence, logical constants, invariance criteria, semantic constraints.

1079-8986/22/2801-0004
DOI :10.1017/bsl.2021.67

of a semantic constraint is: '$I(Red) \cap I(Green) = \emptyset$', which restricts the interpretations of *Red* and *Green* to be mutually exclusive.[1] An exposition of the framework is given in Section 2 and more examples of constraints are given in Section 3.

In the framework of semantic constraints, criteria for logical terms lose much of their significance, as it is semantic constraints of all sorts that determine logical consequence. A semantic clause fixing a term's interpretation is one kind of semantic constraint, while others merely restrict the term's interpretations, or set semantic relations between terms. It is argued that the motivations leading to fixing some terms rather than others don't warrant dismissing other kinds of constraints on interpretations, and so logical consequence should be explicated through the broader framework.

The question of logicality then becomes that of whether there is an appropriate criterion for *semantic constraints* rather than for *logical terms*, that will decide which semantic constraints are acceptable in a logical system. One can still use existing criteria for logicality for the completely fixed terms in this framework—indeed, some terms display a combination of properties that makes them strong candidates for being completely fixed.[2] But still, if the analysis of logical consequence through the broader framework is to be considered, the variety of other constraints needs to be addressed. The option of dismissing out of hand semantic constraints that do not completely fix a term is available, but as a resolution of the issue of logicality it would seem ad hoc.

Indeed, a host of attitudes towards logical terms can be generalized to the framework of semantic constraints. One can promote a criterion demarcating the "logical" semantic constraints, or be skeptical about there being a sharp distinction, taking up, e.g., a relativist or pragmatic attitude. The framework seems to sit well with relativism, but it does not force a relativist position.

This paper takes previous work on the framework forward by considering possible conditions that may be imposed on semantic constraints. I focus on invariance criteria, which I bring to the setting of semantic constraints. I define the notion of *invariance under isomorphisms of semantic constraints* in a way that is conservative with respect to the special case of logical terms (the completely fixed terms in a system). This generalization will then help us gain a better understanding of invariance criteria and their role in model-theoretic semantics.

Criticisms of invariance criteria often lean on counterexamples. Following the generalization, we'll also see that some obviously unpalatable semantic constraints turn out to be invariant under isomorphisms. While in the standard, "term-based" setting, the counterexamples are, in many cases, somewhat contrived, in the framework of semantic constraints, the apparent failure of the criterion is much more obvious, and the examples more wide-ranging and pressing. Nonetheless, we'll see that problematic cases in both

[1] Here we understand *Red* and *Green* to mean *red all over* and *green all over*.

[2] See [3, 27].

settings share the same source. Thus, the generalization will also shed light on the current debate on logicality: we'll be able to see more clearly what invariance conditions may give us, and what they may not.

It will become evident that invariance is not a substantive enough condition for choosing the semantic constraints (or logical terms) that will set the consequence relation for a given language. Invariance conditions are important, however, and their rationale is derived from the idea of modelling, by which the particular identity of the building blocks is insignificant. Indeed, my conclusion will be that invariance conditions are primarily a guide to how the model-theoretic apparatus ought to be used. They do not direct us towards the correct extension of the notion of logical consequence (if such exists). When it comes to choosing the appropriate semantic constraints for a language, a whole other set of considerations must be applied. I thus propose to view invariance conditions as methodological *meta-constraints* both in the framework of semantic constraints and in the term-based setting.

My claims can be summed up thus:

- Invariance under isomorphisms can be naturally generalized to apply to semantic constraints, as revealed by Proposition 4.4 (Section 4).
- Invariance under isomorphisms fails badly as a criterion for logicality of semantic constraints. What is missing is an account of the relation between expressions and their meanings and how it should be factorized in the choice of a logical system. The inadequacy of the criterion is already present in the term-based setting, and now we become better equipped to identify its range and its source.
- While invariance under isomorphisms is not a guide for a choice of a logical system in either framework, it does capture an important feature of model-theoretic semantics, and should be considered as a methodological meta-constraint on logical terms and on semantic constraints.

In Section 2, I briefly present the framework of semantic constraints. In Section 3, I compare fixing terms and fixing constraints, and show how the former can be conceived of as a special case of the latter. In Section 4 I present and discuss the criterion of invariance under isomorphisms for semantic constraints, and in Section 5 I draw some lessons from our results. In Section 6 I propose that invariance criteria should be thought of as *meta-constraints*. Finally, I conclude in Section 7, and indicate the directions I think the study of logicality ought to take. For the generalized criterion, I rely on a somewhat nonstandard notion of isomorphism. In Section 8 I clarify the notion of isomorphism used and how it relates to the standard one. The technical bits throughout the paper all refer to the Tarski–Sher criterion of invariance under *isomorphisms*. But the philosophical discussion applies to other invariance criteria as well. In Section 9 I formulate the invariance criterion for semantic constraints using similarity relations more generally rather than isomorphisms.

§2. Semantic constraints: the framework. The framework of semantic constraints [26] allows for a model-theoretic notion of logical consequence that does not depend on a strict division of the language into logical and nonlogical terms. Semantic constraints are statements in the metalanguage that somehow limit or constrain the possible interpretations of expressions in the object language. We've seen the example constraint '$I(Red) \cap I(Green) = \emptyset$', and we'll see more examples in the next section. The idea of the generalization is the following: whatever motivation (epistemic or other) that led to completely fixing some terms (the logical terms) and letting others (the nonlogical terms) vary in the standard, term-based framework for logical consequence, also allows the option of merely constraining interpretations of terms.[3]

We use the terminology of [26]. A language L consists of *terms*: the primitive expressions, and *phrases*: the meaningful expressions—strings of terms and perhaps auxiliary devices, such as parentheses—that are interpreted in models (not every such string is necessarily a phrase). Every term is assumed to be a phrase; any symbol that is not given an interpretation on its own is considered an auxiliary device. A model $M = \langle D, I \rangle$ for L consists of a non-empty set D (the domain) and an interpretation function I which assigns semantic values to the phrases of L. As in this framework no semantic recursive clauses are assumed in the definition of a model, I is free to assign any value from the set-theoretic hierarchy over $D \cup \{T, F\}$: any concocted set that involves members of the domain and the truth values can be a value of a phrase. It is only by imposing constraints that these values, and models in general, are restricted. Even the division of phrases into semantic categories is not assumed at the outset, and is set using semantic constraints. Note, also, that compositionality is not assumed either. One can impose compositionality as a condition on systems of semantic constraints, and we lay out this option shortly.

A *semantic constraint* is a statement in the metalanguage which includes implicit general quantification over models (on domains and interpretation functions), and which imposes a restriction on the class of models for a language. A precise definition of semantic constraints is to be determined given a metalanguage, which we leave open here to some extent. The metalanguage includes the language of set theory, but can be more comprehensive, according to the needs of interpreting the object language.[4] Given a set of semantic constraints Δ, the Δ-models will be the models abiding by the constraints in Δ. In this framework, all relevant notions are defined with respect to a set of semantic constraints in a semantic manner, even the standardly syntactic ones. We define singular phrases, sentences, and logical consequence:

[3] See also [39, 40] for discussion of various restrictions on admissible models that questions the logical–nonlogical divide.

[4] Some minimal restrictions on the metalanguage ought to be imposed to maintain consistency of the framework; these considerations go beyond our present concerns.

Let Δ be a set of semantic constraints, and let a Δ-*model* be a model abiding by the constraints in Δ.

DEFINITION 2.1 (Singular phrase). A phrase p is a *singular phrase* (w.r.t. Δ) if for every Δ-model $M = \langle D, I \rangle$, $I(p) \in D$.

DEFINITION 2.2 (Sentence). A phrase p is a *sentence* (w.r.t. Δ) if for every Δ-model $M = \langle D, I \rangle$, $I(p) \in \{T, F\}$.[5]

Note that a phrase that in some Δ-models refers to a truth value and in others it does not, will not be considered a sentence (w.r.t. Δ) by this definition.

DEFINITION 2.3 (Logical consequence). Let φ be a sentence and let Γ be a set of sentences. The argument $\langle \Gamma, \varphi \rangle$ is *logically valid* (φ is a *logical consequence* of Γ; w.r.t. Δ) if for every Δ-model $M = \langle D, I \rangle$, if $I(\gamma) = T$ for every $\gamma \in \Gamma$, then $I(\varphi) = T$.

The framework of semantic constraints is a generalization of standard logic: it allows for a formulation of standard first-order logic (including the standard semantic clauses), but also for more or less constrained logics. Some semantic constraints (such as the one above) resemble *meaning postulates* as in the work of Carnap [9] and Montague [21], expressed in the metalanguage.[6] On the other hand, semantic constraints can be taken to represent partial information about a language (e.g., as in '$I(p) \neq I(q)$' which tells us only that p and q always get different interpretations).

Importantly, as opposed to meaning postulates, semantic constraints are not formulated on the backdrop of a system of syntactic and semantic recursive clauses—nor is any division between logical and nonlogical terms assumed at the outset. Indeed, the basic syntactic categories assumed are just those of terms and phrases. One may appeal to an elaborate syntax, based on further syntactic distinctions and suitable recursive clauses—but that would be an addendum to the basic framework. So, for instance, given a set of semantic constraints, we may have a theory of syntax that provides the set of sentences with respect to that set of constraints.

The framework allows for expressions that are fixed in various manners—also with respect to each other. So we obtain various relations between expressions in the language. Let us mention two main definitions from [26]:

DEFINITION 2.4 (Determinateness). A phrase a is *determined by the set of phrases* B (w.r.t. Δ) if for any two Δ-models $M = \langle D, I \rangle$ and $M' = \langle D', I' \rangle$, if $I(b) = I'(b)$ for all $b \in B$ then $I(a) = I'(a)$.

[5]The setting is classical, but surely one can use it to formulate non-classical systems forgoing bivalence.

[6]We note that meaning postulates as in Carnap and Montague are a special case of semantic constraints: the former are sentences formulated in the object language which all admissible models are required to satisfy; the latter are formulated in the metalanguage, and so they are expressively richer.

Definition 2.5 (Dependency). A set of phrases A *depends* on the set of phrases B (w.r.t. Δ) if there are Δ-models $M = \langle D, I\rangle$ and $M' = \langle D, I'\rangle$ sharing the same domain D such that for any Δ-model $M^* = \langle D, I^*\rangle$, if $I^*(b) = I(b)$ for all $b \in B$, then $I^*(a) \neq I'(a)$ for some $a \in A$ (that is, fixing the phrases in B in a certain way excludes some interpretation for the phrases in A that can otherwise be realized).

The notion of dependency is important, as it may serve to highlight the departure from standard, term-based systems, where there are no dependencies as defined above. The present discussion will not make further appeal to this notion. The notion of determinateness will be used extensively, as it is employed in explicating the notion of a *fixed term* that will be central in what follows. It will also be used to define *compositionality*, which is briefly discussed below.

In previous work on the framework the question of whether there is a *correct* system of constraints for logic remained open. Some of the intuitions that seem to be guiding the discussion on logicality of terms are lost here. There do not seem to be any firm intuitions as to which semantic constraints are clearly logical.[7] Indeed, the framework seems to invite a relativistic or pragmatic perspective: each set of semantic constraints yields a logical consequence relation, and different sets might be suitable for different purposes. In this paper I consider a different stance, and take into account conditions that may be imposed on a system of semantic constraints. Ultimately, the appropriateness of a system of constraints, as any system of logic, depends on the use it is intended for. Specifically, a system of semantic constraints can be used in empirical semantics to model logical consequence in natural language, or it can be used as a framework of commitments made by a reasoner [28]. It may be that these two uses (not to exclude other possible uses) will pull us in completely different directions, but it may also be that they do not differ categorically in the conditions they require on constraints, rather only in emphasis and degree: while the former aims at empirical adequacy, the latter might entail a preference for coherence and robustness.

Here, we'll look at conditions on constraints, *meta-constraints*, that would, at least *prima facie*, be relevant to a variety of uses. The rest of the paper will be devoted to invariance conditions. Before that, I shall briefly mention another meta-constraint, that in standard logical systems is taken for granted. I've mentioned that compositionality is not assumed here at the outset. Compositionality, or the idea that the semantic value of a complex expression is a function of the semantic values of its parts and the manner of their composition, can be formally stated in various ways.[8] We define compositionality here as a property of a language with respect to a set of

[7] We note the obvious, that intuitions about logicality are surely affected by (and feed into) what is considered a "standard" logical system, and become ineffective when moving to a less standard or more general setting.

[8] For discussion, see [23].

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press

constraints by reference to all terms occurring in the given phrase, as well as the domain. For the sake of this definition, add to L the term $\mathcal{D}$, and in Δ include the constraint '$I(\mathcal{D}) = D$'.

Definition 2.6 (Compositionality). A language L is *compositional* (w.r.t. Δ) if each phrase p is determined by $\{a : a$ is a term occurring in $p\} \cup \{\mathcal{D}\}$ (w.r.t. Δ).

Example: Assume that *Rudolf*, *sits*, *likes*, and *everything* are terms, and that the following are phrases: *Rudolf sits* and *Rudolf likes everything*. If we employ a set of semantic constraints Δ with respect to which the language is compositional, the semantic value of both phrases in a Δ-model will only depend on the order of the terms they involve, the semantic values of the terms in the given model, and the model's domain.

Some definitions of compositionality require syntactic distinctions beyond those assumed here. If, for example, we added on to our rudimentary syntax a notion of *immediate subexpressions* forming a complex expression (so our strings would come with 'analysis trees' representing their derivation histories; see [23]), then we could define a stronger notion of compositionality, and have every phrase be determined merely by its immediate subphrases and the domain.[9]

Compositionality has been assumed to be a desirable, if not a requisite feature of logical systems—whether they are intended for modelling natural language or as providing an alternative means for reasoning [8, 13]. Whether natural language is in fact compositional (in one sense or another) has nonetheless become a matter of dispute.[10] In the present context, let us simply flag compositionality as an optional meta-constraint in the framework.

§3. Fixing terms vs. fixing constraints. We now have two semantic frameworks at hand: the framework of semantic constraints and the standard, term-based framework. Each framework offers a variety of logical systems that are determined by a choice of semantic constraints or of logical terms that are held fixed. In previous work [26], systems in the term-based semantics were presented as a special case of systems of semantic constraints. Given a set of constraints, the terms of a language can be thought of as fixed in different manners and to different degrees. The completely fixed terms of

[9] Note that the present definition is weak also in that it allows for a compositional language in which the interpretation of a phrase will vary in models with different domains even if the interpretations of its constituent terms are unchanged. The idea is that structural input (the order of the terms and the use of auxiliary symbols) might affect the interpretation of a phrase in a way that is sensitive to the domain (cf. [16, 39] where logical terms are substituted by structural features). Still, the present definition is strong enough for the sake of item 2 in Proposition 9.2 (Section 9) linking elementary equivalence and isomorphism in compositional languages.

[10] For discussion, see [24]. See also [1, 15, 17].

a language with respect to a set of semantic constraints function as the fixed terms in term-based semantics.

What does it mean that we “fix” the logical terms? Usually, what is meant is that we include those terms in the language and assign to them a semantic clause, which says how expressions constructed using a logical term should be interpreted, given the semantic values of their components. In customary presentations of model-theoretic semantics for predicate logic, logical terms are not assigned extensions as such: they are treated as formula building operators that have a respective semantic clause. A term is thus fixed as logical if its semantic action is fixed for all models—if the semantic clause provides an operation by which the interpretation of the complex expression is a function of the interpretations of its components and possibly the domain. Connectives and quantifiers, for example, form well-formed formulas out of well-formed formulas—the semantic value of the well-formed formula formed is a function of the semantic values of the well-formed formulas from which it was formed, and possibly the domain. If we were to adopt this approach in the framework of semantic constraints, then logical terms would not be terms per se, as we assumed that all terms are phrases and are thus interpreted. In that case, logical terms would behave as auxiliary symbols: they would affect the interpretations of phrases in which they appear without having a meaning of their own—in other words, logical terms would be treated as “punctuation marks.”[11]

In the literature on logicality it is customary to assign extensions to logical terms, as various criteria for logicality are applied to a term vis-à-vis its extension. Henceforth I shall follow this route.[12] I thus assume that all terms in the language are assigned an associated operation (usually aiming to capture antecedent use), by which they are tested for their logicality. A term that passes the test can then, as is customary, be completely fixed in the system. In the term-based framework, terms that do not pass the test remain maximally variable—any interpretation that is in accord with their semantic category can be assigned to them in some model (regardless of their associated operation). Now, a term is completely fixed if the interpretation function has no freedom, as it were: given a domain, the interpretation of the term is then determined. So, for example, if we construe quantifiers as second-level predicates in the Fregean fashion, then fixing the standard quantifiers entails that the class of models satisfies the following constraints:

- $I(\forall) = \{D\}$,
- $I(\exists) = \{A \subseteq D : A \neq \emptyset\}$.

[11] Compare the view of logical constants as punctuation marks from the proof-theoretic perspective in [11]. See also the aforementioned [16, 39] where logical terms are substituted by structural features of expressions.

[12] Even when logical terms are only defined through recursive clauses without an extension assigned by the interpretation function, we can look at the set-theoretic operators determined by these clauses and apply the following to them.

Note that the interpretation of the quantifiers, and of the fixed terms in general in term-based semantics, is a function of the domain.[13] More generally, it is assumed that each candidate logical term t has an associated operation O_t which, given a nonempty set D, assigns to t its extension, taken from the set-theoretic hierarchy over $D \cup \{T, F\}$. And so, $O_\forall(D) = \{D\}$. Another example would be $O_{Even}(D) = \{x \in D : x \text{ is an even number}\}$. We do not assume that every term in the language has an associated operation (not every term is assumed to be a candidate logical term or a candidate for fixing), so O_t might not exist for every t. Nor do we assume that O_t, when it exists, is definable in the metalanguage. Presumably, if t is fixed as logical, then it will have an associated operation so that its interpretation in every model will agree with what t's associated operation assigns to the model's domain (see the definition below).

Note that in standard, term-based semantics, when one fixes the interpretation of some terms, one thereby determines a class of models (all those models that give unintended interpretations to those terms are ruled out). Now, by contrast, assume that one obtains a class of models through a set of semantic constraints or in any other way. Given a class of models, one can then draw out the terms that are (completely) fixed in that class as those whose interpretation is a function of the domain. And so we define:

Definition 3.1 (Fixed term). A term t is *fixed* in a given class of models if there is an associated operation O_t such that for every model $M = \langle D, I \rangle$ in the class, $I(t) = O_t(D)$ (i.e., the interpretation of t is a function of the domain).

The definition given here of *fixed term* intentionally does not assume the framework of semantic constraints, but of course one can speak of the fixed terms with respect to a set of constraints that provides the given class of models. We can then loosely say that a term is fixed (w.r.t Δ) if and only if it is determined by the domain. More precisely, including the term $\mathcal{D}$ in L and the constraint '$I(\mathcal{D}) = D$' in Δ as before, we have that a term is fixed in the class of Δ-models if and only if it is determined by $\{\mathcal{D}\}$.[14]

Now, criteria for logicality ordinarily apply to a term t given an associated operation O_t. The idea is that in a correct system for logic, only the terms satisfying the given criterion should be fixed, and the criterion applies to their interpretation were they fixed according to their associated operation. In the next section I adjust this idea to the framework of semantic constraints. We shall look at a constraint C_t that fixes a term t to have an interpretation in accord with an associated operation O_t:

$$C_t : \;\; I(t) = O_t(D)$$

[13] Truth-functional connectives can be construed as functions on truth values or as set-theoretic operations on variable assignment sequences. In either case, the interpretation of the truth-functional connectives is a function of the domain—in the former case, vacuously so, as the interpretation is constant across domains.

[14] See the definition of *logical term* in the "shallow sense" (or *completely fixed term*) in [26].

(read: the interpretation of t in a given model is the value of O_t on the model's domain). Criteria for logical terms can then be reformulated as criteria for the matching constraints. In case that the latter criteria can be naturally extended to apply to all semantic constraints (not just those of the form above), they can be viewed as generalizations of the former. Note, however, that C_t is not a *term*, but it is a *statement*: accepting it does not only entail fixing t, but it is also a statement to the effect that O_t is t's associated operation by which it is fixed. On the other hand, when choosing t as a logical term in the term-based framework, the relation to O_t is normally *presupposed* rather than explicitly asserted. In the term-based framework, the debate is on whether certain terms, say, *there exist infinitely many*, *Unicorn*, etc., are logical, *given* their associated operation. But the *choice* of associated operation is also part of the construction of a logical system, whether implicit or explicit. In the framework of semantic constraints, this choice is made explicit—different operations by which a term can be fixed may be considered through competing semantic constraints, and choosing among them is an explicit stage in the construction of a system in the framework.

§4. Invariance under isomorphisms: terms and constraints. Among criteria for logicality, invariance under isomorphisms is the most prominent in contemporary literature (see [22, 32]). Invariance under permutations too has gained prominent support in the literature [19, 20, 37]; invariance under isomorphisms is a natural way to extend invariance under permutations from single domain to multiple domain model-theoretic semantics. Various criticisms have been raised, mostly accusing isomorphism-invariance of overgenerating. However, many writers on the topic hold that invariance under isomorphisms is at least a necessary condition on logicality, and when a precise definition of logicality is not the main issue discussed, invariance under isomorphisms is the natural default criterion to assume (see [5]).

One of the main driving principles for isomorphism-invariance is that logic ought not be concerned with the particular identity of individuals [4, 22, 32, 36]: however we permute the objects in the domain or switch them with objects from other domains, and the interpretations of the logical terms will maintain their contribution to truth conditions. To make this idea more precise in the present setting, we define the notion of isomorphism and then state the criterion of invariance under isomorphisms for both logical terms and semantic constraints.

We shall say that two models for a language L are isomorphic when there is a bijection from the domain of one to the other which preserves interpretations. To give a precise definition, we need to first extend bijections in the appropriate way to functions that apply to any possible interpretation of phrases in the language. Recall that values of the interpretation function may be any set-theoretic entity constructed from elements of the domain and truth values. We thus define recursively, for any function f from a set D to a set D' the function f^+ on elements in the set-theoretic hierarchy over $D \cup \{T, F\}$ as follows (we assume that $T, F \notin D \cup D'$ and also that D and

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press

D' consist of *ur-elements*, and so they include as members no sets built over D and D', so that a recursive definition can be applied):

- $f^+(x) = \begin{cases} f(x) & \text{if } x \in D, \\ x & \text{if } x \in \{T, F\}, \end{cases}$

and for a set A belonging to the set-theoretic hierarchy over $D \cup \{T, F\}$,

- $f^+(A) = \{f^+(B) : B \in A\}$.

Since this extension of f is a natural one, we shall omit the superscript and simply speak freely of f applying to the relevant sets.[15]

Now we can give the definition for *isomorphic models*, which builds on the definition of *isomorphic models with respect to a set of phrases*:

Definition 4.1 (Isomorphic models). We say that $M = \langle D, I \rangle$ is *isomorphic to $M' = \langle D', I' \rangle$ w.r.t. a set of phrases* S ($M \cong_S M'$) if there is a bijection $f : D \to D'$ which, when appropriately extended, yields $f(I(p)) = I'(p)$ for every phrase $p \in S$. We say that M and M' are *isomorphic* ($M \cong M'$) if they are isomorphic w.r.t. the set of all phrases in the language.

Note that the present notion of isomorphism is stronger than the standard one. Standardly, say, in first-order logic, isomorphism is defined with respect to the nonlogical vocabulary, and then is *proven* to extend to complex expressions in the language. In the present setting, isomorphism with respect to the set of all phrases in the language and isomorphism with respect to the nonlogical vocabulary (whatever that may be) may come apart. We therefore take the stronger definition of isomorphism between models for the formulation of invariance criteria, as the distinction between logical and nonlogical terms no longer plays a central role. I spell out the added assumptions that would entail the equivalence of the weaker and the stronger notions of isomorphism in Section 8.

Now we can formulate invariance criteria for terms and for constraints. I shall first present the definition of invariance under isomorphisms pertaining to terms in the language. As mentioned earlier, when a term t is claimed to be invariant under isomorphisms, it is standardly assumed that there is an operation O_t associated with t that gives its intended interpretation in all domains which is invariant under isomorphisms, and thus we have:

Definition 4.2 (Invariance under isomorphisms: terms). Let t be a term and O_t be the operation associated with t. The term t is *invariant under*

[15]The only seemingly nontrivial choice in the definition was to have f constant on the truth values. The reason for this is that we view the truth values as part of the semantic apparatus as opposed to the material from which domains are made of. When later on we say that invariance under isomorphisms is to capture the thought that logical terms are insensitive to the particular identities of elements in the domain, we do not mean to include the particular identity of True and False, which are very much relevant to logic. This would allow truth-functional connectives to be invariant under isomorphisms by the subsequent definitions.

isomorphisms if for any nonempty sets D and D' and a bijection $f : D \to D'$ appropriately extended, $f(O_t(D)) = O_t(D')$.

The standard logical terms of first-order logic are invariant under isomorphisms. In addition, so are generalized quantifiers such as *Most* and $\exists_{\aleph_0}$ (*there exist infinitely many*). For details, see [31]. For example, we can look at the equality symbol and its associated operation: $O_=(D) = \{< x, x >: x \in D\}$. A bijection from D to D' will invariably take $O_=(D)$ to $O_=(D')$. On the other hand, any term whose operation distinguishes between elements in the domain will not be invariant. For example, the term *Even* taken as a first-level predicate with the associated operation $O_{Even}(D) = \{x \in D : x \text{ is an even number}\}$ is not invariant under isomorphisms, since a bijection might take an even number from one domain to one that is not in another.

Moving on to the proposed generalization, we define what it is for a semantic constraint to be invariant under isomorphisms. Here, again, one might be guided by the motivation that logic should not be sensitive to the individual identity of objects. This motivation can be explicated here as the demand that logic will not distinguish between isomorphic models:

Definition 4.3 (Invariance under isomorphism: semantic constraints). A semantic constraint C is *invariant under isomorphisms* if for any models M and M' such that $M \cong M'$, if M is a $\{C\}$-model, then M' is a $\{C\}$-model.[16]

Note that the models M and M' in the definition are *any* models, not necessarily satisfying some given set of constraints.

We shall see a variety of constraints that are and others that fail to be invariant under isomorphisms shortly, but first let us consider one sort of example to help us see the relation between the two notions of invariance. We consider, for each term t which has an associated operation O_t, its associated semantic constraint which fixes t with the operation O_t, as was formulated in Section 3:

$$C_t : \quad I(t) = O_t(D).$$

Indeed, such constraints allow us to reformulate term-based systems in the framework of semantic constraints. For each term fixed as logical in a term-based system, we can include its matching constraint, ultimately leading to the same class of models (assuming other constraints pertaining to more general semantic and syntactic features of the logic are also included, e.g., setting the semantic categories of the non-fixed terminology). And so, for example, we can consider the constraint

$$C_= : \quad I(=) = O_=(D).$$

[16]Compare [38, p. 790], where Zimmermann formulates the "meta-constraint" that any appropriate class of models for model-theoretic semantics is closed under isomorphisms. We return to Zimmermann in Section 6.

It is easy to see that this constraint on models is closed under isomorphisms, whereas

$$C_{Even}: \quad I(Even) = O_{Even}(D),$$

is not.

By the definition of invariance under isomorphisms for terms, we immediately get that if t is invariant under isomorphisms, then every two $\{C_t\}$-models are isomorphic w.r.t. $\{t\}$. Now, if t is invariant under isomorphisms, it does not distinguish between equinumerous models, and thus its relevant constraint should be invariant under isomorphisms. In fact, the entailment goes both ways.

Proposition 4.4 *Let t be a term that has an associated operation O_t, and let C_t be its associated constraint. Then t is invariant under isomorphisms iff C_t is invariant under isomorphisms.*

Proof. For the left to right direction, assume that t is invariant under isomorphisms. Let $M = \langle D, I\rangle$ and $M' = \langle D', I'\rangle$ be isomorphic models, and assume that M is a $\{C_t\}$-model. So $I(t) = O_t(D)$. By the assumption that t is invariant under isomorphisms, $f(O_t(D)) = O_t(D')$. So $f(I(t)) = O_t(D')$. By the assumption that $M \cong M$, $f(I(t)) = I'(t)$, and so we have $I'(t) = O_t(D')$ proving that M' is $\{C_t\}$-model, and so that C_t is invariant under isomorphisms.

For the right to left direction, assume that C_t is invariant under isomorphisms, D and D' are nonempty sets, and $f : D \to D'$ is a bijection. In order to make use of our assumption, we employ isomorphic models with the given domains. Let $M = \langle D, I\rangle$ and $M' = \langle D', I'\rangle$ be models with the given domains and with interpretation functions that satisfy: (a) $I(t) = O_t(D)$ and (b) for every phrase p, $I'(p) = f(I(p))$. (There are such models: recall that our definition of models is very liberal, and we can let, e.g., $I(p) = \emptyset$ for every phrase $p \neq t$.)

By (a), M is a $\{C_t\}$-model, and by (b), $M \cong M'$. By our assumption that C_t is invariant under isomorphisms, M' is a $\{C_t\}$-model. So $f(O_t(D)) = f(I(t)) = I'(t) = O_t(D')$ as required. ⊣

I thus contend that the definition of invariance under isomorphisms for semantic constraints can be appropriately regarded as a generalization of the definition for terms. The thought is as follows. As applied to terms, invariance under isomorphisms means that the term is indifferent to distinctions between elements of the domain, it is blind to switching and exchanging elements in the domain: an isomorphism-invariant term does not distinguish between isomorphic models. And this is precisely the property which we use to define isomorphism invariant semantic constraints. Indeed, Proposition 4.4 is an indication that the definition caught on to the desired property. If the move to the more general framework of semantic constraints for the explication of logical consequence is accepted, then our definition gives the appropriate corresponding generalization of isomorphism-invariance.

We can now observe some consequences of the generalized definition. From the examples below it immediately becomes apparent how the criterion of invariance under isomorphisms pertains only to the form of constraints, and it is completely indifferent to the intended meaning of the terms that appear in those constraints. The following constraints are invariant under isomorphisms:

1. $I(\forall) = \{D\}$.[17]
2. $I(Red) \cap I(Green) = \emptyset$.
3. $I(Even) \cap I(Odd) = \emptyset$.
4. $I(Bachelor) \cap I(Married) = \emptyset$.
5. $I(\mathbf{3}) \in I(Odd)$.[18]
6. $I(Unicorn) = \emptyset$.
7. $I(Water) = I(H_2O)$.
8. $|I(Red)| \geq 375$ (i.e., the size of the extension of *Red* is at least 375.)

If we take the predicates in the constraints above to have meanings associated with them that are derived from natural language correlates, then it seems that the possible grounds for accepting the above constraints would be of different nature. While the grounds for the first five constraints would presumably be a-priori: conceptual and perhaps also metaphysical or mathematical, those for the last three would be a-posteriori if not straightforwardly empirical and, at least in the last constraint, contingent, if such grounds would even exist. Of course, the criterion allows constraints that patently defy intended meaning, either as a matter of mathematical fact:

9. $I(Even) = I(Odd)$

or as a matter of empirical fact:

10. $I(Red) \cap I(Big) = \emptyset$.

The criterion of invariance under isomorphisms thus seems to be quite permissive. However, our proof above shows that in the case of the fixed terms, it is completely aligned with the standard invariance criterion. And the generalized criterion does rule out some constraints—those which refer to specific objects in domains, such as:

11. $I(naturalNumber) = \{0, 1, 2, ...\}$.
12. $3 \in I(Odd)$.
13. $I(Even) \cap I(Prime) = \{2\}$.

So the generalized criterion, although weak, is not completely powerless in ruling out semantic constraints. To be sure, in some sense the generalized criterion captures the thought that the particular identities of elements in

[17] Recall that semantic constraints implicitly generalize over domains and interpretation functions: I stands for an interpretation function, D stands for a domain.

[18] I use '**3**' (in boldface) to stand for a term in the language, and '3' to stand for the number three.

models should not matter. By this we mean that it does not matter what kind of object each element of the domain is; what material it is made of, so to speak. On the other hand, the *structure* that the interpretation function imposes on the domain is not something the invariance criterion will have a say on, where the *structure* includes any set-theoretic relation as well as properties of sets having to do with cardinality. By the criterion, we can consider as logical whatever meaning relations we like, as long as the material of the domain is not constrained. In the following sections we make further observations on the forgoing generalization.

Before we move on, let us look at another class of constraints that is sanctioned by the invariance criterion:

14. $I(\mathbf{5} + \mathbf{7} = \mathbf{12}) = T$.
15. $I(Bachelor(John)) = T$.
16. $I(\mathbf{0} = \mathbf{1}) = T$.

In general, for any phrase p, the semantic constraint '$I(p) = T$' is invariant under isomorphisms: this is a trivial outcome, given our robust notion of isomorphism. Note that even on a weaker notion of isomorphism such constraints would turn out to be invariant, if we add some standard assumptions to the framework that lead from isomorphism to elementary equivalence (see Section 8). And thus, constraints imposed by meaning postulates or theories in the object language are invariant under isomorphisms: for any theory $\mathcal{T}$, there is a set of constraints Δ that are invariant under isomorphisms such that the Δ-models are the models satisfying $\mathcal{T}$ (see also [38, p. 790]).

This outcome may be used as a further rebuttal of invariance under isomorphisms as a criterion for logicality of semantic constraints. Still, the thought that model-theoretic semantics should be indifferent to the particular identities of elements in a model is preserved. Indeed, we may note that theories in the object-language can never distinguish between particular members of the domain, as they never have direct access to members of the domain that is not mediated by language. This is the gist of Putnam's permutation argument against model-theoretic accounts of reference [25]: reference can only be fixed up to isomorphism. It is of interest to note that in the present context of discussion, this feature is perceived favourably: from the perspective of logicality, particular members of the domain should not be distinguishable.[19] This leaves us with a condition that cannot possibly fulfil the role of a *criterion* for the acceptance of semantic constraints into a system, but can still serve for an initial screening. In the next section I analyse the outcomes of the generalization, and in the subsequent section I re-evaluate the role of invariance in setting up a logical system.

§5. Lessons from the generalized criterion. The criterion of invariance under isomorphisms for logical terms has been widely discussed and analysed in philosophical literature. Nonetheless, the discussion suffers from

[19]See [33, 34] for discussion on the contrast between logical concerns and the concerns of a theory of reference.

some blind spots. The generalized viewpoint offered here can further the understanding of the philosophical role of invariance conditions in standard systems as well as in systems of semantic constraints. Let us go over several immediate lessons from looking at the isomorphism invariance criterion for semantic constraints. In the next section I will use these lessons in formulating the role I think invariance conditions should play in logical systems.

I. *Vocabulary*. The first lesson has to do with the relation of invariance under isomorphisms to vocabulary that is empirical or refers to contingent properties. The term 'red' (by its intended meaning) is not normally associated with an operation that is invariant under isomorphisms, and thus it fails the criterion for logical terms. Indeed, completely fixing the interpretation of 'red', if that is at all possible,[20] would require us to divide any given domain into the extension and the anti-extension of 'red', and assuming there are cases where neither of them is empty, we shall have to make distinctions between particular objects. However, merely *constraining* 'red' does not necessarily commit us to these distinctions. The constraint '$I(Red) \cap I(Green) = \emptyset$' does not refer to any particular objects: whether or not it holds will be so even if elements are switched and moved around. The first lesson from the generalization to semantic constraints is that even if we have a principle, such as isomorphism invariance, prohibiting us from completely fixing some term, it does not follow that the interpretation of the term needs to remain maximally variable. We might still allow the term to be constrained in various ways, abiding by the principle in its generalized form. Moreover, invariance criteria, as understood here, do not produce a bifurcation of the vocabulary into its "formal" and its "non-formal" parts: all terms can be fixed at least to some extent and thus take part in a "logical" language. Invariance criteria rather tell us which ways of fixing meanings or meaning-relations are admissible.

II. *Logicality vs. fixity*. Proponents of invariance criteria pose a twofold demand on logical terms of a system: that they have a fixed interpretation, and that this interpretation is invariant under isomorphisms (or other transformations). Whatever is not completely fixed cannot be logical. The second lesson is that logicality and fixity can be pulled apart, even under the assumption that invariance under isomorphisms is a test for logicality (if we are open to generalizing it to the proffered framework). While a term without an associated operation completely fixing it cannot even be considered for logicality in the term-based framework, the generalization of invariance criteria allows for clauses or rules that merely limit a term's interpretation.

[20]In [27] I argue that isomorphism invariance is a pre-requisite for fixing a term in a manner faithful to its meaning in standard extensional model-theoretic semantics.

As an example from recent literature, we can look at the study of Carnap's categoricity problem of deriving semantic interpretation of terms from their associated inference rules. In a recent article, Bonnay and Westerståhl [6] extend the discussion from propositional to first-order logic, and analyse the categoricity problem of the quantifiers. The rules for the universal quantifier do not fix its interpretation completely according to its usual semantic clause, but rather narrow down the admissible models to those where the interpretation of the universal quantifier is a principal filter closed under the interpretation of the singular terms in the model. Absence of categoricity will appear to be troublesome to those who think that we learn the interpretations of the logical constants via their inference rules. In their analysis, Bonnay and Westerståhl take for granted that logicality entails fixity. They then show that if the universal quantifier is also required to be invariant under permutations, then together with the inference rules one obtains a fixed meaning. And thus, by imposing a commonly accepted condition for logicality, one can obtain a fixed logical term.

Note that the restriction on models that is obtained by Bonnay and Westerståhl, by the inference rules for the universal quantifier, can be formulated as a set of semantic constraints. The restriction to models where the universal quantifier is interpreted as a principal filter generated from some subset A of the domain is formulated by the semantic constraint: '$I(\forall) \in \{\{B \subseteq D : A \subseteq B\} : A \subseteq D\}$'. Closure under the interpretation of the singular terms in the model is captured by including the constraint '$I(t) \in \bigcap\{B : B \in I(\forall)\}$' for each singular term t in the language.[21] These semantic constraints are invariant under isomorphisms. Thus, the present contribution to that discussion will be: even though the rules for the universal quantifier do not suffice for fixing the meaning of the term, they do constrain it in a *logical* way—that is if isomorphism-invariance is endorsed as a criterion for logicality (Bonnay and Westerståhl themselves propose permutation-invariance as a condition for topic-neutrality). What the rules fail to provide are *fixed* terms. Logicality (of relevant constraints), under the common assumptions, is granted.

Note, also, that rules formulated in the object language will always yield isomorphism-invariant constraints: this is for the same reason that was mentioned at the end of the previous section concerning meaning postulates and theories in the object language.

[21]Recall that *singular terms* in the framework of semantic constraints are terms whose interpretation is a member of the domain in each model (admissible by a given set of constraints). Of course, the category of singular terms in this framework coincides with that of singular terms in the special case of standard systems, and so the constraints above are another way of stating Bonnay and Westerståhl's results. Note that by the completeness of first-order logic, the addition of "non-standard" interpretations does not make any difference to the extension of the relation of logical consequence.

III. *Modelling*. Invariance under isomorphisms is part and parcel of practice in model-theoretic semantics.[22] The more general moral that can be drawn from the foregoing is that by the invariance criterion, the models of model theory are really just *models*: different "materials" or "substances", so to speak, can be used to model the same phenomenon. Passing the invariance criterion's test for logicality simply means that this feature of modelling is maintained in the semantic framework. Invariance is thus a desirable property if models are to be used. As it happens, if you restrict your perspective to completely fixed terms, and apply invariance criteria to this class of terms, it looks like a bifurcation of the vocabulary is at issue. However, when viewing things more broadly, considering all sorts of semantic constraints, a strict division of the vocabulary into the logical and the nonlogical is no longer the main outcome. The issue rather seems to be the basic features of being a model, of what kinds of restrictions on models *qua* models are permitted.[23]

To sum up, invariance criteria, when generalized, do not bifurcate the language, in the sense that they do not sieve the empirical, contingent, or non-formal expressions out of logical languages. Constraints on all parts of vocabulary can be invariant. Further, invariance criteria, in their generalized form, do not require a strict division of the language into the fixed and non-fixed terms: terms that are not completely fixed can have constraints that are yet invariant under isomorphisms. Finally, invariance criteria pertain to the way in which the semantic apparatus, namely the models, are to be used: distinctions between isomorphic models have no place in model-theoretic semantics. If a certain meaning relation formed as a semantic constraint is not invariant under isomorphisms, this tells us that the standard model-theoretic apparatus is inadequate for capturing it. From this we can derive that the meaning relation is too complex, or indeed non-logical; but it would be good to remember that we arrived at these properties through a certain semantic apparatus for modelling.

Nonetheless, defenders of invariance criteria might claim that in the special case of the term-based framework, invariance under isomorphisms (or other transformations) *does* do the work of dividing the language in a more or less desirable manner, getting the distinction between logical and nonlogical terms approximately right—assuming, of course, the pairing of candidate terms with their associated operations. I agree that in the term-based setting, invariance criteria do not fail as badly, but I think that they leave much to be desired. To be sure, terms that are invariant under

[22] Generalized quantifiers in [22] are invariant under isomorphisms by definition. Closure under isomorphisms of the class of models for a logic is assumed throughout [2].

[23] Shapiro [30] distinguishes between two kinds of features of models: representors—those elements in models that correspond to some features of the phenomena modelled, and artefacts—those elements in models that do not represent, but are part the model, perhaps making it simple or coherent. Here I take invariance criteria to express the idea that the material from which elements of the domain of a model are made is an artefact.

isomorphisms have associated operations that are simple and general, as they don't distinguish between equinumerous domains. Nevertheless, the association of terms with operations, which is assumed in the background of term-based systems, and is made explicit in the framework of semantic constraints, is a crucial factor in the adequacy of a logical system. In the next section, I stress the importance of this element in the construction of a logical system, on which invariance criteria are silent.

§6. Invariance criteria as meta-constraints. The following arguments are valid in systems of semantic constraints that are invariant under isomorphisms (including the constraints '$I(Bachelor) \cap I(Married) = \emptyset$' and '$I(Unicorn) = \emptyset$' as well as the standard constraints for first-order logic):

(A)

$$\frac{Bachelor(John)}{\neg Married(John)}$$

(B)

$$\frac{}{\neg Unicorn(John)}$$

Argument (B) is already captured by the term-based framework abiding by invariance under isomorphisms, assuming that the extension of *Unicorn* is the empty set in every model. Should these arguments be sanctioned by a system for logic? Examples such as (B) can be used to cast doubt on invariance under isomorphisms as a criterion for logicality, and indeed, some have used *Unicorn* as a counterexample for this criterion.[24] The present discussion of invariance criteria from a general perspective can help us locate the issue with these putative counterexamples. It seems (at least prima facie) wrong to fix *Unicorn* in a logical system, not because its associated operation is somehow nonlogical, but because *the choice and assignment of the associated operation seems to be too contentious to be allowed to have an effect on logical consequence.*

I do not wish to take a definite stance here with respect to (A) and (B) and whether a system for logic should sanction these arguments. But the status of these examples cannot be settled merely by invariance conditions, as those will also sanction:

(C)

$$\frac{Bachelor(John)}{Married(John)}$$

[24] See [14].

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press

(D)

$\neg Man(John)$

We need only make sure that we include the relevant, invariant, semantic constraints (i.e., '$I(Bachelor) = I(Married)$' and '$I(Man) = \emptyset$' as well as the standard constraints for first-order logic). Clearly, what's wrong with these constraints has to do with their relation to the intended meaning of the terms. The Tarskian approach to logical consequence assumes that the (correct) semantic facts are given, and that there's no doubt about them. When choosing a logical system, such assumptions need to be negotiated, and invariance conditions contribute nothing on this matter.[25] Given that the role assigned to criteria for logicality in the relevant discussion is to complete Tarski's definition of logical consequence, invariance criteria fail their purpose badly.

At this point we can mention two possible objections to the critical assessment above, in defence of invariance criteria as adjudicating logicality. The first opts for the use of a fully symbolic language that divorces expressions from their preconceived meanings: while a use of an expression in natural language could lead to a choice of an associated operation, in the formalized setting meaning is reduced to that operation—its source is disregarded, as are all other aspects of meaning. And so, if indeed our semantics happens to dictate that *Unicorn* denotes the empty set in every model, then it means the same as the co-extensional 'non-self-identical' for all intents and purposes of the formal system.[26] From this perspective, what invariance criteria presumably fail to capture is irrelevant to the construction of a fully formal logical system.

The second objection is less radical. Here one accepts that the assignment of an associated operation to an expression might be a non-trivial matter, but still they take it as a given on the basis of which we choose a system for logic. To justify this, one might argue for a division of labour between logical and linguistic or other inquiries. Logic, by this defense, is not supposed to deal with the basis for assigning one meaning or another to a given term, but rather it is there to differentiate between terms given the meanings that they have. Give the logician a language with its semantics, and they will

[25] See also MacFarlane on permutation invariance: "Its main shortcoming is that it operates at the level of reference rather than the level of sense; it looks at the logical operations expressed by the constants, but not at their meanings. An adequate criterion, one might therefore expect, would operate at the level of sense, perhaps attending to the way we grasp the meanings of logical constants" [18]. I agree with the gist of MacFarlane's remark, although I would add that (given that criteria for logicality apply to linguistic expressions), it is the relation between an expression and its meaning that is at stake, and to this relation I refer in what follows. Senses may be involved (I do not go into this topic), but note that (Fregean) senses, while intermediate between a linguistic expression and its extension, are on their own meanings detached from linguistic expressions.

[26] For such a line of argument, see [34, p. 304].

identify its logic. Further, a logician might claim that the subject matter of logic is not language itself, but the formal operations denoted by logical terms. Such a logician might appeal to Tarski [37] for support, where Tarski focuses on logical notions, which he defines as the set-theoretic entities that are invariant under permutations of the domain. One can then define logical terms to be the linguistic expressions that happen to denote logical notions, but logic is primarily concerned with the denotations, not with how language attaches itself to them.[27]

To respond to these objections, I'd like to first reiterate my acknowledgement that logical systems have different uses, serving different purposes giving rise to different conceptions of logic. In a setting where logic is viewed as a purely mathematical discipline (as in [37]), invariance under isomorphisms might indeed suffice as a criterion for logicality. Nonetheless, I contend that in a setting where the validity of meaningful arguments is concerned, the relation between expressions and their formal explications needs to be considered.[28] Criteria for logical terms are intended to fill in the lacuna in defining the concept of logical consequence or of logical validity (see [4, 31]). Logical consequence in the model-theoretic tradition is a relation between linguistic entities. In other words, the question at the base of the project is *which arguments are valid*. We may look at denotations to answer this question. But we cannot disregard the nature of the relation between expressions and their denotations, lest we accept *Unicorn* and other such terms as logical terms—being merely guided by invariance criteria. In the framework of semantic constraints, we would end up with '$I(Water) = I(H_2O)$' and other semantic constraints that are based on empirical findings, which would presumably be unpalatable for anyone who is interested in setting bounds for logic.

Many of the critics of the isomorphism-invariance criterion agree that it serves at least as a *necessary* condition for logicality, but perhaps not as a sufficient one. However, the problem is not that the criterion misses certain cases. The picture under which the isomorphism-invariance criterion pretty much gets it right and is just in need of some modification is misleading in view of its utter silence on fundamental semantic and metasemantic issues concerning the relation between terms and their intended meanings.[29] The kind of overgeneration we encounter, that arises in the stage of meaning assignment, will not be overcome by stricter invariance criteria (see Section 9). Invariance criteria on their own just don't seem to deliver what is expected

[27]Indeed, it is customary to interpret Tarski's later work as completing the explication of logical consequence he set out on in his earlier paper [36], but see [29] for an alternative interpretation.

[28]Note that when presenting the criterion of permutation invariance for logicality, Tarski expressly disconnects the discussion from the issue of logical truth, and logical consequence or validity is not even mentioned [37].

[29]For the distinction between semantics and metasemantics see [7]. A related distinction is made by Stalnaker [35] between *descriptive semantics* and *foundational semantics*. In the former one provides the semantic values of expressions in the language, and in the latter one asks what makes it the case that a language has the descriptive semantics that it has.

of them in terms of guiding us towards a correct logical system, at least when we consider terms with a preconceived meaning coming from natural language.

I suggest to simply give up the phrase "criteria for logicality" when it comes to invariance conditions. Invariance certainly has a role in model-theoretic semantics, but it is not that of marking a line between logical vocabulary and nonlogical vocabulary, or between the vocabulary that should be fixed and the vocabulary that shouldn't be fixed. Nor does it serve as an arbitrator for logical semantic constraints. I suggest, instead, an alternative rephrasing that I adopt from Zimmermann [38]. Zimmermann considers meaning postulates and other kinds of constraints on models as means for approximating appropriate models for English. The method of verification of a correct set of constraints would presumably be any of the existing methods in linguistic theory. Zimmermann then proposes a meta-constraint on classes of appropriate models that they should be closed under isomorphisms, and this is equivalent to our criterion of invariance under isomorphisms for semantic constraints. The meta-constraint does not tell us which classes of models are appropriate—whether a given meaning postulate such as 'No bachelor is married' is correct for English. The meta-constraint is more of a feature of the model-theoretic apparatus. The exact same holds in our setting: the invariance criterion for semantic constraints is not faithful to intended meaning and therefore may not be a guide for the correct logical system—it merely indicates the ways models can be used, and should thus be considered as a methodological meta-constraint.[30]

Restricting our perspective back to term-based logic: we have learned that the criterion of invariance under isomorphisms for logical terms is a special case of the requirement of closure under isomorphisms of classes of models, and thus it too, by Zimmermann's own lights, should be considered as a meta-constraint. It tells us what kinds of operations can be fixed as denotations of terms in the language, without telling us further how and which to fix. To be sure, every criterion for logicality would be a meta-constraint, but not every meta-constraint is a criterion for logicality.

Although Zimmermann is concerned with model-theoretic semantics from a linguistic perspective, this outlook on invariance applies more generally—whether we are concerned with natural language as an empirical object of study or with explicating the concept of logical consequence from the reasoner's perspective. The observations in Section 5 have led there to the claim that invariance criteria tell us how models ought to be used, that the "material" from which the domain is made is an *artefact* and should not affect logical consequence. This sits well with the *invariance as a methodological meta-constraint* perspective. Adopting this perspective means that (a) the critics of invariance *criteria* are correct—invariance falls

[30]Perhaps we can identify a similar attitude in Shapiro [30], who states the *isomorphism property*, a condition that is equivalent to isomorphism-invariance, without endorsing it as a full criterion: "I would submit that the isomorphism property is a necessary condition of any model theory worthy of the name" [30, p. 152].

short of bringing Tarski's explication of logical consequence to fruition, and that (b) inasmuch as invariance tells us anything about logicality, it is through its methodological role in setting-up the boundaries of the model-theoretic framework. Model theory is a tool we use for theorizing in logic, and invariance under isomorphisms should reasonably be included in any manual for the use of this tool (on the assumption that the particular identities of elements of domains should not be considered in modelling).

§7. Conclusion. The aim of this paper was twofold. First, I generalized invariance criteria so as to apply in the framework of semantic constraints. I have done so by reformulating invariance under isomorphisms (of a term) as closure under isomorphisms of classes of models (satisfying a constraint). Secondly, through the generalization, we obtain a new perspective on the philosophical significance of invariance criteria. On the assumption that the move to the more general framework of semantic constraints for the explication of logical consequence is a viable one, the two main results were:

1. Invariance criteria, as generalized, do not provide a strict division of language into two types of vocabulary, but they rather delineate the acceptable ways in which meaning can be presented in model-theoretic semantics. The framework of semantic constraints lets go of the strict distinction between "fixed" and "nonfixed"; we see here that invariance criteria for logicality can be generalized to this framework and therefore have no inherent demand for such a distinction.
2. Invariance criteria are silent on the association of terms and operations, and more generally on the relation between terms and their meanings. This claim is recognizable already in the term-based setting, but its impact is emphasized considerably in the generalized framework.

From 1 and 2 we come to the position that invariance criteria tell us how to use model-theoretic semantics: models present meanings through the structure imposed by the interpretation function. Models merely *model*: the material used in their domains is thus irrelevant; invariance under isomorphisms ensures this ground rule. We learn that the ground rules for using model-theoretic semantics are moot in difficult cases where intended meaning is not trivial, or where matching a term with its intended meaning appears to go beyond what we'd admit as logical considerations. All these together suggest that invariance criteria do not provide us with the desired means for dividing up a language into its logical and nonlogical sections, and that they rather serve as *methodological meta-constraints* on the proper use of model-theoretic semantics.

There may be a lot to learn about a term through the ways its meaning can be represented in model-theoretic semantics. Thus, I do not mean to dismiss the significance of invariance as a tool for studying logicality. I do, however, propose a shift of focus in the discussion, based on the following two reflections: First, whatever invariance criteria teach us, it is through their role as meta-constraints in model-theoretic semantics—we primarily learn something about model-theoretic semantics, and in combination with further assumptions about meaning, we might draw more fundamental conclusions on the logicality of terms or constraints. Secondly and relatedly, the assignment of intended meaning, usually taken for granted, should be considered more thoroughly and seriously in the study of logical consequence. Where do intended meanings come from? How exactly are they represented in model theory—what do models represent, and how do semantic clauses capture meaning? Hopefully, we have provided sufficient motivation to give these issues their proper due.

§8. Appendix A. Isomorphism and elementary equivalence. In this section I discuss the definition of isomorphic models, which, as noted, is stronger than the standard one. To be sure, the present definition of invariance under isomorphisms for terms is consistent with the literature (it does not even refer to the definition of isomorphism, which was employed in the case of semantic constraints). Still, to clarify the relevant notions, we relate the present definitions to standard results. For the standard results we need compositionality and that the operations forming complex expressions are invariant under isomorphisms.

Definition 9.1 (Elementary equivalence). Let $M = \langle D, I\rangle$ and $M' = \langle D', I'\rangle$ be models for L. We say that M and M' are *elementary equivalent* ($M \equiv M'$) if for every phrase p, $I(p) = T$ iff $I'(p) = T$ and $I(p) = F$ iff $I'(p) = F$.

Proposition 9.2

1. *Let M and M' be models. If $M \cong M'$ then $M \equiv M'$.*
2. *Let Δ be a set of semantic constraints. Assume that L is compositional w.r.t. Δ, all fixed terms are isomorphism-invariant, and all the operations interpreting complex phrases according to Δ*[31] *are isomorphism-invariant. Then if M and M' are Δ-models that are isomorphic with respect to the non-fixed vocabulary with respect to Δ, then $M \cong M'$.*

We remark that (1) is immediate and (2) is trivial. Moreover, the implication from isomorphism w.r.t. the nonlogical vocabulary to elementary equivalence in standard first-order logic is a special case of this proposition. Note that as a result, any constraint of the form '$I(p) = T$' for some phrase p will be invariant under isomorphisms (see the discussion at the end of Section 4).

[31] Let L be compositional w.r.t. Δ, and let p be a phrase that includes the terms $t_1, \dots, t_n$ (in that order) with perhaps auxiliary symbols. Then there is an $(n+1)$-place operation O_p^Δ such that for every Δ-model $M = \langle D, I\rangle$, $I(p) = O_p^\Delta(D, I(t_1), \dots, I(t_n))$. We call O_p^Δ the operation interpreting p in L according to Δ. We say that O_p^Δ is invariant under isomorphisms if for any sets D and D' and a bijection $f : D \to D'$, $f(O_p^\Delta(D, I(t_1), \dots, I(t_n))) = O_p^\Delta(D', f(I(t_1)), \dots, f(I(t_n)))$.

§9. Appendix B. Other invariance criteria. The criterion of invariance under isomorphisms as applied to logical terms has been accused of overgenerating, based on various kinds of counterexamples other than those we've been considering. First, there are terms whose meaning varies with the cardinality of the domain. A quantifier $Q_{\exists\backslash\forall}$ that is interpreted as the existential quantifier on domains of the size of a successor cardinal and as the universal quantifier otherwise is invariant under isomorphisms, but because it doesn't seem to have a unified meaning is deemed by many as intuitively nonlogical.

In addition, all the cardinality quantifiers are invariant under isomorphisms, including $\exists_{2^{\aleph_0}}$ (*there exist continuum many*) and $\exists_{\aleph_1}$ (*there exist* $\aleph_1$ *many*). The equality of the extensions of these quantifiers is independent of the ZFC axioms of set theory, and so it seems highly undesirable to many to permit them as logical terms.

The literature on logicality contains various accounts that modify the isomorphism-invariance criterion. One line of proposals takes instead of isomorphisms, a wider class of transformations, so that the class of invariant terms becomes more exclusive. Two such proposals will be presented and extended to the framework of semantic constraints. We shall see that indeed, some semantic constraints will be ruled out by these proposals. However, the lessons drawn in Section 5 will apply to these proposals just as well.

Thus, instead of invariance under isomorphisms, we can consider more generally invariance under similarity relations, as in [4, 12]. Here, we take a similarity relation to simply be a relation between models without further conditions (but some conditions can be imposed if needed). Isomorphism, as defined in the previous section, is a similarity relation.

Definition 10.1 (Invariance under similarity relations: semantic constraints). Let $\mathcal{S}$ be a similarity relation between models. We say that a constraint C is *invariant under* $\mathcal{S}$ if for any models M and M', if M is a $\{C\}$-model and $M\mathcal{S}M'$, then M' is a $\{C\}$-model.[32]

Now we need to phrase the condition of invariance under an arbitrary similarity relation with respect to terms. Since the similarity relation may not necessarily be determined by a class of functions, we need a more general treatment than the one we had for isomorphisms in Section 4. The definition we give is in line with the specific cases of invariance under a class of functions, and is indeed very close to some formulations in the literature.

Definition 10.2 (Invariance under similarity relations: terms). Let $\mathcal{S}$ be a similarity relation between models, t a term, and O_t its associated operation. We say that t is *invariant under* $\mathcal{S}$ if for any models M and M', if $I(t) = O_t(D)$ and $M\mathcal{S}M'$ then $I'(t) = O_t(D')$.[33]

[32]Simply put: the class of models satisfying the relevant constraint is closed under $\mathcal{S}$.

[33]To show that the definition is in line with standard usage, we compare it to Bonnay's definition of S-invariance of operators. Bonnay deals with quantifiers as second-level predicates, and structures of the form $\langle M, A\rangle$ where M is a domain and A a subset of that domain. Operators associated with quantifiers are then functions on such structures,

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press

From this definition, the equivalence of invariance under a similarity relation of a term t and of its associated constraint C_t falls out immediately. But this is no surprise, given the discussion (and proof) in Section 4.

Let us consider two cases of similarity relations: homomorphisms and potential isomorphisms.

DEFINITION 10.3 (Homomorphic models). We say that $M = \langle D, I \rangle$ is *homomorphic* to $M' = \langle D', I' \rangle$ ($M \mathcal{H}om M'$) if there is a surjection $f : D \rightarrow D'$ such that, when properly extended, gives for every phrase p: $f(I(p)) = I'(p)$.

Feferman has endorsed a variant of invariance under homomorphisms as a criterion for logicality.[34] Applied to terms, we get the standard quantifiers ($\forall$ and $\exists$), but we don't have the identity relation or any of the non-standard cardinality quantifiers ($Q_{\exists \setminus \forall}$, $\exists_{2^{\aleph_0}}$ and $\exists_{\aleph_1}$, and also $\exists_{\aleph_0}$, $\exists_{\leq 3}$, etc.). According to Feferman, his criterion ensures homogenous meaning and absoluteness with respect to models of set theory.

As for semantic constraints, on the present definition of homomorphisms, the following examples pass the criterion. We note that there may be other, close ways of defining homomorphisms which would yield different results.[35]

- $I(Even) \cap I(Prime) \neq \emptyset$.
- $I(Unicorn) = \emptyset$.
- $I(Big) = I(Green)$.
- $I(Blue) \subseteq I(Extended)$.
- $|I(Red)| \leq 375$.

The relation of homomorphism is non-symmetric, and the property of invariance under homomorphisms is not closed under negation of constraints (or complement, if a constraint is identified with a class of models). The following semantic constraints are not invariant under homomorphisms:

- $I(Blue) \cap I(Green) = \emptyset$.
- $I(Big) \neq I(Green)$.
- $|I(Red)| = 375$.
- $|I(Red)| \geq 375$.

assigning T to structures where the second element falls under the quantifier's extension in the structure's domain and F otherwise. Bonnay then defines: "We say that an operator Q is $\mathcal{S}$-invariant iff, for any structures $\mathcal{M}$, $\mathcal{M}'$, if $\mathcal{M}\mathcal{S}\mathcal{M}'$, then $Q(\mathcal{M})$ iff $Q(\mathcal{M}')$" [4, p. 39]. Now, $Q(\mathcal{M})$ is a short way from saying that $\mathcal{M}$ fixes the quantifier according to its associated operation.

[34]The criterion he ultimately proposes (in [12]) is not simply invariance under homomorphisms, but rather the following, more restrictive one, which does not so easily render itself to generalization to the framework of semantic constraints: an operation is logical if and only if it is λ-definable from monadic homomorphism-invariant operations.

[35]For the delicacy of the criterion of invariance under homomorphisms to slight variations in framework, see [10].

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press

We make two observations with respect to invariance under homomorphisms as a criterion for semantic constraints. First, it seems that the boundary delineated is rather arbitrary: is there a reason to distinguish between the relation of being mutually exclusive and its complement? In addition, while constraints setting the exact cardinality or a lower bound on the cardinality of a term's extension are excluded, constraints setting an upper bound are not.

Secondly, we see that the same phenomena observed in Section 6 with respect to the criterion of invariance under isomorphism can be observed here: the points regarding *Unicorn* hold just the same.[36] The same will apply to the criterion of invariance under potential isomorphisms, due to Bonnay [4], to which we turn next, in advance of further elaboration of this point.

Definition 10.4 (Potentially isomorphic models). We say that $M = \langle D, I \rangle$ and $M' = \langle D', I' \rangle$ are *potentially isomorphic* ($M \cong_p M$) if there is a non-empty set of *partial isomorphisms* $\mathcal{F}$ from D to D', meaning that for every $f \in \mathcal{F}$:

- for every $d \in D$, there is a $g \in \mathcal{F}$ such that $f \subseteq g$ and $d \in dom(g)$, and for every $d' \in D'$, there is a $g \in \mathcal{F}$ such that $f \subseteq g$ and $d' \in ran(g)$,
- when appropriately extended, we have for every phrase p, $f(I(p) \cap dom(f)) = I'(p) \cap ran(f)$.

Bonnay's criterion for logical terms too blocks some counterexamples to isomorphism-invariance: we have both absoluteness and some assurance of homogenous meaning (so $Q_{\exists\setminus\forall}$, $\exists_{2^{\aleph_0}}$ and $\exists_{\aleph_1}$ are all blocked, though not $\exists_{\aleph_0}$, $\exists_{\leq 3}$, for which the intuitions are less clear). More generally, potential isomorphisms distinguish between finite cardinalities, but not between infinite ones.

The semantic constraints invariant under potential isomorphisms include all those having to do with set-theoretic relations such as being mutually exclusive and subset inclusion, as well as their complements. As for "cardinality-constraints", similar to the case of terms, we have constraints having to do with countable (finite and infinite) cardinalities (e.g., '$|I(Red)| = 375$') but not those involving uncountable cardinalities (e.g., '$I(realNumber) = 2^{\aleph_0}$').

It can easily be observed that even though some of the putative counterexamples to the isomorphism-invariance criterion for logical terms are blocked, the lessons we drew in the previous sections still apply. The alternative invariance conditions can be adopted as meta-constraints as suggested with respect to isomorphism invariance—an option I shall not discuss here.

[36]Interestingly, if we look at a term such as $maleWidow$, understood as 'a male that is a widow' we observe that here '$I(maleWidow) = \emptyset$' and '$I(Male) \cap I(Widow) = \emptyset$' come apart: the former is invariant under homomorphisms and the latter isn't.

§10. Acknowledgments. Material from this paper was presented in Bergen, Jerusalem, Munich, and at the 2016 Annual Meeting of the Association for Symbolic Logic—I thank the audiences there for comments and feedback. I'd also like to thank Eli Dresner, Ran Lanzet, Sam Lebens, Hannes Leitgeb, Oron Shagrir, Stewart Shapiro, and Jack Woods for comments and discussion. This research was supported by the ISRAEL SCIENCE FOUNDATION (grant no. 1954/17).

REFERENCES

[1] N. Asher, ***Lexical Meaningin Context: A Web of Words***, Cambridge University Press, Cambridge, 2011.

[2] J. Barwise and S. Feferman, editors, ***Model-Theoretic Logics***, Springer, New York, 1985.

[3] J. van Benthem, *Logical constants across varying types*. ***Notre Dame Journal of Formal Logic***, vol. 30 (1989), no. 3, pp. 315–342.

[4] D. Bonnay, *Logicality and invariance*, this Journal, vol. 14 (2008), no. 1, pp. 29–68.

[5] ———, *Logical constants, or how to use invariance in order to complete the explication of logical consequence*. ***Philosophy Compass***, vol. 9 (2014), no. 1, pp. 54–65.

[6] D. Bonnay and D. Westerståhl, *Compositionality solves Carnap's problem*. ***Erkenntnis***, vol. 81 (2016), pp. 721–739.

[7] A. Burgess and B. Sherman, ***Metasemantics: New Essays on the Foundations of Meaning***, Oxford University Press, Oxford, 2014.

[8] R. Carnap, ***Meaning and Necessity***, The University of Chicago Press, Chicago, 1947.

[9] ———, *Meaning postulates*. ***Philosophical Studies***, vol. 3 (1952), no. 5, pp. 65–73.

[10] E. Casanovas, *Logical operations and invariance*. ***Journal of Philosophical Logic***, vol. 36 (2007), no. 1, pp. 33–60.

[11] K. Došen, *Logical constants as punctuation marks*. ***Notre Dame Journal of Formal Logic***, vol. 30 (1989), no. 3, pp. 362–381.

[12] S. Feferman, *Logic, logics and logicism*. ***Notre Dame Journal of Formal Logic***, vol. 40 (1999), no. 1, pp. 31–55.

[13] G. Frege, *Sense and reference*. ***The Philosophical Review***, vol. 57 (1948), no. 3, pp. 209–230.

[14] M. Gómez-Torrente, *The problem of logical constants*, this Journal, vol. 8 (2002), no. 1, pp. 1–37.

[15] J. Groenendijk and M. Stokhof, *Why compositionality*?, ***Reference and Quantification: The Partee Effect*** (G. Carlson and J. Pelletier, editors), CSLI Press, Stanford, 2005, pp. 83–106.

[16] T. G. Haze, *The accident of logical constants*. ***Thought: A Journal of Philosophy***, vol. 9 (2020), no. 1, pp. 34–42.

[17] T. M. Janssen and B. H. Partee, *Compositionality*, ***Handbook of Logic and Language*** (J. van Benthem and A. T. Meulen, editors), Elsevier, Amsterdam, 1997, pp. 417–473.

[18] J. MacFarlane, *Logical constants*, ***The Stanford Encyclopedia of Philosophy (Fall 2009 edition)*** (E. N. Zalta, editor), 2009. Available at http://plato.stanford.edu/archives/fall2009/entries/logical-constants/.

[19] F. Mautner, *An extension of Klein's Erlanger program: Logic as invariant-theory*. ***American Journal of Mathematics***, vol. 68 (1946), no. 3, pp. 345–384.

[20] V. McGee, *Logical operations*. ***Journal of Philosophical Logic***, vol. 25 (1996), pp. 567–580.

[21] R. Montague, *The proper treatment of quantification in ordinary English*, ***Formal Philosophy*** (R. Thomason, editor), Yale University Press, New Haven, 1974, pp. 247–270.

[22] A. Mostowski, *On a generalization of quantifiers*. ***Fundamenta Mathematicae***, vol. 44 (1957), no. 1, pp. 12–36.

[23] P. Pagin and D. Westerståhl, *Compositionality I: Definitions and variants*. ***Philosophy Compass***, vol. 5 (2010a), no. 3, pp. 250–264.

[24] ———, *Compositionality II: Arguments and problems*. ***Philosophy Compass***, vol. 5 (2010b), no. 3, pp. 265–282.

[25] H. Putnam, ***Reason, Truth and History***, ***vol. 3***, Cambridge University Press, Cambridge, 1981.

[26] G. Sagi, *Formality in logic: From logical terms to semantic constraints*. ***Logique et Analyse***, vol. 57 (2014), no. 227, pp. 259–276.

[27] ———, *Logicality and meaning*. ***The Review of Symbolic Logic***, vol. 11 (2018), no. 1, pp. 133–159.

[28] ———, *Logic in natural language: Commitments and constraints*. ***Disputatio***, vol. 12 (2020), no. 58, pp. 377–408.

[29] ———, *Extensionality and logicality*. ***Synthese***, vol. 198 (2021), pp. 1095–1119.

[30] S. Shapiro, *Logical consequence: Models and modality*, ***The Philosophy of Mathematics Today*** (M. Schirn, editor), Oxford University Press, Oxford, 1998, pp. 131–156.

[31] G. Sher, ***The Bounds of Logic: A Generalized Viewpoint***, MIT Press, Cambridge, 1991.

[32] ———, *Did Tarski commit "Tarski's fallacy"*? ***The Journal of Symbolic Logic***, vol. 61 (1996), no. 2, pp. 653–686.

[33] ———, *The logical roots of indeterminacy*, ***Between Logic and Intuition: Essays in Honor of Charles Parsons*** (G. Sher and R. Tieszen, editors), Cambridge University Press, Cambridge, 2000, pp. 100–123.

[34] ———, ***Epistemic Friction***, Oxford University Press, New York, 2016.

[35] R. C. Stalnaker, *Reference and necessity*, ***A Companion to the Philosophy of Language*** (C. Wright and B. Hale, editors), Blackwell, Malden, 1997.

[36] A. Tarski, *On the concept of logical consequence*, ***Logic, Semantics, Metamathematics*** (J. Corcoran, editor), Hackett, Indianapolis, 1983, pp. 409–420.

[37] ———, *What are logical notions*? ***History and Philosophy of Logic***, vol. 7 (1986), no. 2, pp. 143–154.

[38] T. E. Zimmermann, *Model-theoretic semantics*, ***Semantics: An International Handbook of Natural Language Meaning*** (C. Maienborn, K. von Heusinger, and P. Portner, editors). Walter de Gruyter, Berlin, 2011.

[39] A. Zinke, ***The Metaphysics of Logical Consequence***, Klostermann, Frankfurt, 2018.

[40] ———, *Reinterpreting logic*, ***The Semantic Conception of Logic: Essays on Consequence, Invariance, and Meaning*** (G. Sagi and J. Woods, editors), Cambridge University Press, Cambridge, 2021.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF HAIFA
199 ABA KHOUSHY AVE
MT. CARMEL, HAIFA 3498838, ISRAEL
E-mail: gilisagi@gmail.com

https://doi.org/10.1017/bsl.2021.67 Published online by Cambridge University Press