THE REVIEW OF SYMBOLIC LOGIC
Volume 11, Number 1, March 2018

LOGICALITY AND MEANING

GIL SAGI
Department of Philosophy, University of Haifa

Abstract. In standard model-theoretic semantics, the meaning of logical terms is said to be fixed in the system while that of nonlogical terms remains variable. Much effort has been devoted to characterizing logical terms, those terms that should be fixed, but little has been said on their role in logical systems: on what fixing their meaning precisely amounts to. My proposal is that when a term is considered logical in model theory, what gets fixed is its *intension* rather than its extension. I provide a rigorous way of spelling out this idea, and show that it leads to a graded account of logicality: the less structure a term requires in order for its intension to be fixed, the more logical it is. Finally, I focus on the class of terms that are invariant under isomorphisms, as they render themselves more easily to mathematical treatment. I propose a mathematical measure for the logicality of such terms based on their associated Löwenheim numbers.

§1. Introduction. Standard formal languages distinguish between logical and nonlogical terminology. By the word *term* I shall refer to a primitive expression that is interpreted in models, be it a quantifier, a predicate or an individual constant. A term treated as logical functions in a certain way: its meaning is said to be held fixed across models. The question at hand is: What does fixing a term's meaning in model-theoretic semantics precisely amount to? This seems to be an issue that has not been properly dealt with. The literature is filled with technically elaborate and philosophically insightful debates on logicality. Thus, while the seemingly deeper question of which terms are or should be considered as logical has received much attention, characterizing their function in formal systems in precise terms has heretofore been neglected. However, any account of logical terms that disregards their role in systems of logic is incomplete. Moreover, the notion of a fixed term is of interest even if it is denied that there is a principled criterion for logical terms. For example, on relativistic or pragmatic accounts of logical terms there are still terms in each particular system which get fixed.

The question raised in this paper concerns a basic philosophical aspect of model theory. Besides the connection to the debate on logical terms, there is the more general issue of how the machinery of models relates to meaning and related philosophical concepts. I show that defining the notion of a fixed term in a system is not a trivial matter, even on the most natural assumptions. I present the problem in §2. Next, I move on to consider the way "meaning" should be understood (§3). Standard first order logic, on which I focus, is said to be a "logic of extension". Since under common ways of assigning extensions to the quantifiers, their extensions are not constant across models but are a function of the domain, we might say that they are not completely fixed. My proposal is that it is the *intension*

Received: October 27, 2016.
2010 *Mathematics Subject Classification*: 03A05.
Keywords and phrases: logical constants, Löwenheim numbers, invariance criteria, model theory, meaning.

doi:10.1017/S1755020317000247

rather than the extension of logical terms that is fixed. On this view, the quantifiers are completely fixed vis-à-vis their intension. Intensions will be understood through possible worlds. In §4, I relate models to the philosophical apparatus of possible worlds, and in §5, I define what it is to fix a term in model theory.

We then reach an interesting result. In §6, I will show that the assumptions made in previous sections entail that all terms that are fixed according to a preconceived meaning are invariant under isomorphisms. The meaning of terms that are not isomorphism-invariant is too complex to be fixed in standard systems: it requires more structure than ordinary models, consisting of just a domain and interpretation function, can provide. Terms that are not isomorphism-invariant are thus precluded from being considered logical terms even before a criterion for logicality enters the picture, because they are incapable of performing the assigned role in formal systems. However, as I will show in §7, adding some structure to formal systems makes it possible to fix the recalcitrant terms.

The addition of structure required for fixing some terms can be seen as compromising their logicality. A connection between logicality and the amount of structure required in order to fix a term will thus present itself. I will not propose a new criterion for logicality, but rather a scale: the less structure that is required to fix a term in a manner faithful to its intended meaning, the more logical it is. When we come to the question of which terms should be fixed, the scale provides one point of consideration among others.

In the resulting scale, invariance under isomorphisms provides a dividing line between those terms that can be faithfully fixed in standard model theory and those that cannot. And so, §8 will deal with the former kind of terms—and to those I propose an absolute measure for logicality, using the model-theoretic notion of a Löwenheim number. Those terms that carry with them a lower Löwenheim number require less structure in order to be fixed, and are thus more logical.

The upshot is that discussions on logicality need to account for the way that meaning is incorporated in a logical system. This task is tackled here in a model-theoretic framework. The outcome is shown to bear significant implications on the question of criteria for logical terms—those terms whose meanings get fixed. These implications are spelled out through a graded account of logicality.

§2. The problem. In standard model-theoretic logical systems, we fix, by means of the interpretation function, the meaning of the connectives and the quantifiers, and perhaps the identity relation. Other terms can presumably be fixed as well. The fixed terms receive their own special symbol and a metalinguistic semantic clause. So, we have a clause fixing the interpretation of '$\wedge$' and '$\forall$', but only a general grammatical categorization of an arbitrary predicate 'P'. For any candidate fixed term we thus assume that there is an associated interpretation that is either fixed or not. This interpretation is somehow derived from the meaning of a natural language counterpart, or may simply be introduced in the system.

Recent decades have seen a productive discussion on logical terms in model-theoretic semantics. *Invariance under isomorphisms* is probably the most widely accepted criterion for logical terms (see, e.g., Sher (1991, 1996), McGee (1996)). There have been accusations of both overgeneration and undergeneration, leading to alternative invariance accounts (Feferman, 1999; Bonnay, 2008; Woods, 2014). Other kinds of criteria have been proposed as well, taking into account basic features of logic such as its generality, formality, necessity and apriority (see, e.g., Peacocke (1976), McCarthy (1981), Hanson (1997), Warmbrōd (1999), Gómez-Torrente (2002)). Some, on the other hand, have cast doubt on the existence of a principled distinction between logical and nonlogical terms that could serve as a basis for an account of logical consequence (van Benthem, 1989; Read, 1994; Dutilh Novaes, 2012).

It is a common view, at least among those who support a principled distinction between logical and nonlogical terms, that only logical terms should be fixed in a correct system for pure logic. The focus in the literature has been on finding criteria for logical terms—those terms that should be fixed in a system—while an explanation of their role in being fixed is largely missing.[1] When we do look at their role in formal systems, two questions arise. The first is formulated by Ken Warmbrōd as follows: "The question which must be addressed by any criterion of logical constancy, then, is this: why should those terms that satisfy the criterion, and only those terms, have a fixed meaning in a theory of logical truth and consequence?" (Warmbrōd, 1999, p. 505). This question is especially significant if logical consequence or logical truth are the primary notions we are interested in in the construction of formal systems. Why does fixing the logical terms yield a good formal account of the more primary notions?

Secondly, before answering the above question, we should also ask: What is it to fix a term in a system? This question has virtually been ignored in the literature. To begin with, there is the technical notion of fixing a term in model theory, which deserves a proper, precise definition—beyond the description opening this section. Then, there is the more philosophical notion of fixing a term in accordance with its intended meaning. While for the initial, technical notion we need not make reference to intricate philosophical concepts, the latter requires some philosophical machinery.

It is usually simply taken for granted that one can fix the meaning of a term in model-theoretic semantics. Nonetheless, model theory is a mathematical tool, so the question of how such a tool can capture the meaning of some terms of a given vocabulary is nontrivial. In this paper, I will mainly focus on the second question from the perspective of standard contemporary model theory. The question of what is a fixed term is relevant not only for those who contend that there is a principled distinction between logical and nonlogical terms, but anyone working with model-theoretic semantics and claiming that some terms are fixed. Dealing with the question of what is a fixed term will then lead us to new observations on logicality and the value of fixing some terms rather than others, and so we shall also make progress with respect to Warmbrōd's initial question.

Before moving on to an account of fixity in contemporary model theory, we can mention two precursors: substitutional semantics and single domain model-theoretic semantics. In both these settings the notion of fixity is straightforward. Let us examine them in turn. For simplicity, let us focus on logical truth rather than logical consequence. In substitutional semantics, commonly attributed to Bolzano (1929), the logical truth of a sentence is determined by the truth of a range of sentences that share the same form, that are substitution instances of the original sentence. A substitution instance of a sentence is obtained from the original one by substituting nonlogical terms for nonlogical terms of the same syntactic category in a uniform manner. In substitutional semantics, a sentence is logically true if and only if all its substitution instances are true. So, for example, we can derive that "All

[1] There is some equivocation in the use of the phrase "logical terms" in the literature. Sometimes, what is meant by "logical terms" is simply "those terms that are fixed in a system": this is a technical, system-relative use. Otherwise, the phrase is attributed a more philosophical sense, referring to terms that exhibit some feature which renders them logical, regardless of whether they are actually fixed in a given system. To avoid confusion, I use the phrase "fixed terms" for the first sense, and reserve the use of "logical terms" for the second.

humans are mortal" is not a logical truth by substituting "sitting" for "mortal". There is no use of models here, just the material truth of a range of sentences. The logical terms are fixed here in the sense that they are not substituted—we might say this is a straightforward notion of syntactic fixity.

Single domain model theory is just like ordinary contemporary model theory, only there is just one domain—the universe of all existing objects. In the case of single domain model theory, we have a Tarskian reduction of logical truth to material truth. All nonlogical terms in the sentence are substituted by variables of appropriate type in a uniform manner, and universal quantifiers binding all new variables are attached to the sentence. The original sentence is said to be a logical truth if and only if the resulting sentence is materially true (Tarski, 1936, pp. 416f). Here we can derive that "All humans are mortal" is not a logical truth by looking, for example, at the assignment function that assigns the class of sitting things to the (second level) variable substituted for "mortal". Here the logical terms are again those that are not substituted, and so are syntactically fixed. Furthermore, if we consider the logical terms in single domain semantics to have extensions (as we'll do in what follows), we can say that logical terms in single domain semantics are completely fixed in the sense that they have the same extension in all models. This would then entail a semantic notion of having a fixed interpretation.

In contemporary model theory, where multiple domains are considered, no terms are substituted: a logical truth is a sentence that is true in all models—under all reinterpretations. Here, to see that "All humans are mortal" is not a logical truth, we give an interpretation where the extension of "humans" does not fall under the extension of "mortal". Syntactic fixity is thus irrelevant: the sentence does not undergo any syntactic manipulation. Neither do we have semantic fixity in the sense of a fixed extension, as noted in the following quote from Sher. Here Sher describes the difference in the ways logical and nonlogical terms are treated in contemporary model theory:

> The nonlogical terms are strongly variable: any formally possible denotation in accordance with their formal "skeleton" (i.e., their being individual terms, n-place relation of individuals, etc.) is represented in some model. The logical terms, on the other hand, denote fixed formal properties of objects, and their denotations are subject to the laws governing these properties. *These terms are fixed not in the sense that they denote the same entity in each model* (the denotation of the universal quantifier in a model with 10 elements differs from its denotation in a model with 11 elements). Rather, while nonlogical terms are defined within models, logical terms are defined by fixed functions over models. (Sher, 1996, p. 675, my emphasis)

The next three sections will elaborate on this quote. We shall discuss the extensions of logical terms focusing mainly on quantifiers, and explain how logical terms may be fixed even if their extension is not completely fixed.

§3. Fixing the meaning of logical terms: extension or intension? The function of logical terms in formal systems is commonly described as that of being fixed, that is, the *meaning* of logical terms is said to be fixed in the semantics. Nonlogical terms, by contrast, are those whose meaning varies.[2] In the restricted contexts we shall allude to (leaving out

[2] See McCarthy (1981, p. 499), Warmbrōd (1999, p. 505).

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

context sensitivity and hyperintensionality), the meaning of a term can be understood as either its *extension* or its *intension*. Since it is extensional logic we are dealing with here, it would make sense to look at extensions first. The standard logical terms of first-order logic are usually presented through recursive clauses for truth-conditions. However, they can also be assigned explicit extensions. Thus, for instance, the quantifiers can be defined in a Fregean manner as second level predicates (see, e.g., Sher (1991, p. 10)): For a model $\langle D, I \rangle$, with domain D and interpretation function I, the interpretation of the quantifiers is as follows:

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

Note that this way of treating quantifiers opens the door to a wide range of *generalized quantifiers* defined as second level predicates. Notable examples are *Most* and *There are infinitely many*.[3]

We can see in the examples above that the interpretation of the quantifiers is relative to the domain: their extension, although constrained by the above rules, is not constant: it is not *completely* fixed. The situation with the truth-functional connectives is different. Those can be defined extensionally as denoting truth functions and are not affected by changes in the domain. For example:

- $I(\neg) = f_\neg$ where $f_\neg$ is a function from truth values to truth values such that for a truth value v, $f_\neg(v) = T$ iff $v = F$.
- $I(\wedge) = f_\wedge$ where $f_\wedge$ is a function from pairs of truth values to truth values such that for truth values v_1 and v_2, $f_\wedge(v_1, v_2) = T$ iff $v_1 = v_2 = T$.

By contrast to both the quantifiers and the truth-functional connectives, the standard nonlogical terms—predicates and individual constants—may receive in a model any extension agreeing with their semantic type, and in that sense they are completely variable, or *un*fixed. It thus may be suggested that fixity is a matter of degree (Kuhn, 1981). This is certainly true if we look at the fixity of *extensions*. Nonetheless, there is a sense in which quantifiers *are* fixed, and is not simply that they are nonpurely-variable. Given a domain, there is only one possible extension for each of the quantifiers. This property is common to all logical terms in standard extensional systems. '∀' does not have the same extension in all models, but it always means "all objects" or "all of the domain". Likewise with '∃' and '='. As I will show later on, there is a difference between connectives and quantifiers with respect to *what it takes* to fix their meaning, that is, how much structure they require, and there we do find varying degrees.

I thus propose that it is the *intension* rather than the *extension* of a logical term that is fixed. Talk of intensions is not out of place in this setting, even though it is standard extensional logic we are concerned with. Let an intension of a term be a function from possible worlds to relevant extensions in those worlds, according to the semantic type of the term. In order to accommodate intensions in this setting, we shall look at models (still in extensional logic) as representing possible worlds. Only when a connection between the mathematical apparatus and the philosophical concepts is made, can we claim to fix the *meaning* of a term in a system.

[3] An early characterization of generalized quantifiers appears in Mostowski (1957). For an extensive presentation of the topic, see e.g., Westerståhl (1989).

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

§4. Models and possible worlds. Indeed, the correlation between models and possible worlds is especially appealing if we endorse a principle of *necessity* for logical validity, saying that in a logically valid argument, the conclusion follows necessarily from the premises. If we accept that the necessity involved in the principle is metaphysical, and that metaphysical necessity can be spoken of in terms of possible worlds, then we have the following condition on validity of arguments:

> *Necessity*: In all possible worlds in which the premises are true, so is the conclusion.

Many authors take *Necessity* on board, and this imposes a demand that there be some connection made between models and possible worlds.[4]

The way of connecting models and possible worlds I will adopt, is that of viewing models as *representing* or *modeling* possible worlds, in accordance with Stewart Shapiro's logic-as-model approach (Shapiro, 1998). Model theory as a whole can be seen as a mathematical discipline that can be used to explicate the notion of logical consequence by way of modeling. In this way, models are conceived of as pure mathematical entities (constituted by a pair of a domain and an interpretation function). Alternatively, there is a common practice of taking domains of models themselves to contain concrete things. There are strong reasons to prefer the modeling approach, and as we shall see later on (§6), having concrete objects as the members of model-domains is unhelpful with respect to the issues at hand. At this stage, we can note as a reason for adopting the modeling way that when models are kept "mathematically pure", they are not tangled in all the complications of possible worlds, at least when standing on their own. Thus we have mathematical formalism on the one side, employed by philosophical theory on the other.[5]

Now, there are different ways models can be seen as representing possible worlds. First, we have Etchemendy's *representational semantics* (Etchemendy, 1990, 2006). According to representational semantics, each possible world, or each way the world could be is represented by some model. Representational semantics is one of two contrasting perspectives on model theory described by Etchemendy, the other being *interpretational semantics*, where models are considered as reinterpretations of the language, not taking into account alternative metaphysical possibilities. Interpretational semantics is considered by many not to be a viable option, precisely because it does not take modality into account, and so does not ensure the satisfaction of *Necessity*. Etchemendy attributes the interpretational approach to Tarski—whether he is correct in doing so is a matter of dispute (see Gómez-Torrente (1996)). On the other hand, representational semantics has been shown by Etchemendy and others (Sher, 1996) to be a poor account of model-theoretic consequence, partly because employing models would entail all the metaphysical complications of possible worlds.

However, representational semantics is not the only theory by which models represent possible worlds. We also have Shapiro's blended approach—blending the representational

[4] From here onwards I opt for *possible worlds* rather than *situations* or *states of affairs*. This choice is inconsequential in what follows—which should be chosen is not our current concern. What is crucial is that we have a range of items in which expressions receive their extensions (serving as the domain of intensions), and truth in which accounts for the modal aspect—the necessity—of logical consequence.

[5] See also Zimmermann (1999, 2011), who advocates the mathematical purity of models in the context of natural language semantics.

and the interpretational perspectives—by which models represent possible worlds under reinterpretation of the nonlogical vocabulary (Shapiro, 1998). We assume that formal language sentences have natural language correlates, the former are true or false in models, the latter in possible worlds. On this approach, if a sentence φ of a formal language is true in a model M, then M represents a possible world w under some interpretation of the nonlogical vocabulary of a natural language correlate of φ, s, such that s holds in w under that interpretation. One of the basic ideas of the blended approach is that there need not be a one-one relation between models and possible worlds, and so not every metaphysical question needs to be addressed. Models take into account additional parameters, such as reinterpretation of the nonlogical vocabulary. The reinterpretation of the nonlogical vocabulary is presumed to ensure the formality of logical consequence or validity. For *Necessity* we merely need that every possible world be represented by some model—that there is agreement between truth in a model and truth in the possible worlds represented by that model under some reinterpretation. Arguably, the blended approach provides a good account of model-theoretic consequence (see also Sagi (2014)), and this is the approach I shall adopt hereafter.

One might worry that even though models remain purely mathematical and generally metaphysically innocent, having them represent possible worlds brings in more metaphysical complications than logic can countenance, just as in the case of representational semantics. However, since we let the interpretation of nonlogical terms vary, the said complications can largely be avoided—if we make the right choice of logical terms (Shapiro, 1998; Sagi, 2014).

It would be natural, furthermore, to view elements of models as representing objects inhabiting possible worlds, and more generally, domains of models as representing populations of possible worlds.[6] Arguably, the following is a natural assumption to make in model theory:

> *Replacement*: A domain D represents a population of a possible world w if and only if there is a one-one and onto correspondence between members of D and objects in w. Each such correspondence constitutes a relation of representation, where each member of D represents an object in w.[7]

In particular, a model's domain represents a population of a possible world if and only if they are of the same cardinality. This assumption sits well with model-theoretic practice, where a countable domain can be used to represent the natural numbers, the integers or the rational numbers. The interpretation function, which forms with the domain

[6] I shall use "populations" when referring to domains of possible worlds, to be distinguished from domains of models.

[7] See e.g., an appeal to *Replacement* in the presentation of model-theoretic semantics in Zimmermann (2011). Since we are dealing with standard model theory here, all domains of models are restricted to be sets. We thus assume that all domains of possible worlds are set-sized, and this can be viewed as problematic (given that there are good reasons to think of the domains of some possible worlds as proper-class-sized). I will nonetheless set aside these complications and make this assumption without getting into the details of justifying it. Also, for simplicity, we shall add the assumption that there are possible worlds of all cardinalities, so that each model represents at least one possible world. Later on, I shall note explicitly when this assumption should be reassessed.

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

a model, narrows down which structure is represented in the end. We shall see later that this assumption has significant consequences, but let us keep it for now.[8]

Once we have set the relation between models and possible worlds in place, we can employ it to accommodate intensions. But first, we should say how set-theoretic constructs over the domain of a model represent complex extensions over possible worlds. So, for instance, a subset A of a domain D represents a set B of objects from a possible world w under some correspondence between D and w if and only if A consists of all and only elements of D representing objects in B on that correspondence. More generally, Let f be a one-one and onto correspondence from D to $pop(w)$, where '$pop(w)$' designates the population of w. Let us extend f to apply to set-theoretic constructs over D and $pop(w)$ by recursion in the natural way. So, for a set A in the set-theoretic hierarchy constructed over D, $f(A) = \{f(a) : a \in A\}$.[9]

We now know how to use set-theoretic constructs to represent the values of intensions over possible worlds. When a term is fixed according to its intended meaning (when it is *fixed faithfully*) the values it gets by the interpretation function should correspond to the values of its associated intension. However, note that we have not defined what kind of construct represents an intension—it cannot be a set-theoretic construct over one domain—it must be over many, as it concerns a range of possible worlds. In the next section I provide the definitions of *fixed* and of *fixed faithfully*, and on the way we shall see what is used to represent an intension.

§5. What is a fixed term? An intension, we have assumed, is a function from possible worlds to extensions therein. Intensions will thus be represented by functions from models to extensions therein. Note that until now, when speaking of representation of possible worlds by models, I only considered domains of models, not interpretation functions. The reason is as follows. The interpretation function is defined on the basis of a choice of fixed terms. Recall the discussion in §3: the interpretation function assigns to the fixed (logical) terms their fixed interpretation, and to the nonfixed (nonlogical) terms any extension consistent with their semantic category. The interpretation function assumes a set of terms that are to be fixed (at least in the usual case), and indeed—it presumably fixes those terms by assigning to them an interpretation that conforms to their meaning. For this reason, the interpretation function itself cannot be an argument for a function representing an intension: the interpretation function will be defined according to such functions. Therefore, the domain will be the only parameter we shall employ when fixing terms in a given model with that domain (at least in the case of standard models—consisting of just a domain and interpretation function).

We are now in a position to give a proper response to the main question set forth. We can distinguish between two relevant notions. First we have a more technical notion of a fixed term. Here we define this notion for standard model-theoretic semantics, where we have a (nonempty) set as domain and besides that just an interpretation function. We shall thus say that a term is *fixed* in standard model-theoretic semantics if its extension is a function of the domain, namely, it is constant across models sharing the same domain. Thus, for each fixed term t, we have an operation O_t from domains to extensions such that for any model

[8] We should note another, more basic assumption that we are making here: we have a picture of possible worlds as inhabited by objects that have properties and relations, as opposed to possible worlds as constituted by, e.g., propositions. We shall keep this assumption throughout.

[9] To keep things simple, we require domains to be composed of ur-elements—enabling the recursion to work by blocking a set from being both a member and a subset of the domain.

$M = \langle D, I\rangle$, $I(t) = O_t(D)$. For this notion we need not make reference to intensions or possible worlds. The standard quantifiers are obviously fixed in standard model theory by this definition. Note also that the symbols for truth-functional connectives in standard model theory satisfy the definition: their extensions are constant across all models (the same functions on truth values in each model) and are thus a (constant) function of the domain. More generally, this definition is in accord with standard practice in model theory: the fixed (logical) terms are normally given recursive semantic clauses stating how they should be interpreted, and those semantic clauses employ the domain as a parameter by which the logical term is interpreted.

Now, in many cases we are not merely interested in fixing a term, but we also wish to capture a preconceived meaning. Note that a fixed term carries with it a function from domains to extensions which is precisely the type of thing that can be used to represent intensions. We want this function to give us the right extension in every model—that whatever that extension represents in some possible world would be the value of the intension in that possible world. We shall thus say that a term associated with an intension is *fixed in a manner faithful to its intended meaning*, for short: *fixed faithfully* if it is fixed, and the value of the associated operation in every model represents the value of the intension in every possible world represented by the model (on some correspondence), and no other extension. In other words, for a term to be faithfully fixed, its extension in each model must not represent anything but its extension in some possible world. When a term is fixed faithfully, we shall say that the associated operation represents the associated intension.

Note that faithful fixing requires taking into account all available relations of representation. In principle, we could have restricted ourselves to a limited class of correspondences acting as representations and defined faithful fixing accordingly. This option will be mentioned later on, and as we shall see, such a restriction may carry with it metaphysical complications we might like to avoid. For now, note that *Replacement* has it that any one-one and onto correspondence counts as a relation of representation.

In this setting, one might inquire about the conditions for successfully fixing a term in a manner faithful to its meaning—can any intension be captured in this way, considering the wide range of correspondences? The next section will present a condition on intensions that can be thus represented, and so on terms that can be faithfully fixed. If we want to be able to faithfully fix other terms we shall have to change our model theory or modify *Replacement*. For now, note that if discussions on logicality are to relate at all to terms that have some preconceived meaning, the notion of having them fixed faithfully is crucial. This means that their associated operation when fixed is in accord with their meaning: our way of representing the intension naturally suggests itself.

Now, we have noted that the interpretation of logical terms is somehow restricted, even if not constant. In standard logic, '$\forall$' is as fixed as can be: its intension is not a constant function, so it makes sense that its interpretation will not be constant across domains. Indeed, we may view the operation $O_\forall$ denoted by '$\forall$' as representing the associated intension in the class of models, the first giving to every domain D the extension $\{D\}$ as value (see §3), and the latter giving to every possible world w the extension $\{pop(w)\}$ as value (where '$pop(w)$' denotes the population of w). The idea I am suggesting is that the extension of a fixed term is allowed to vary: the variation in extensions should correspond to the variation in the values of its associated intension.

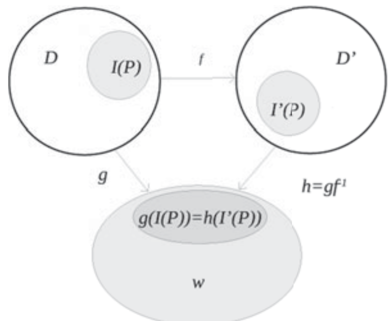
The above definitions can be made more general, so as to apply to other, less standard settings: either where the models are of a different structure, or where the assumptions relating them to possible worlds are altered. Such examples will be given in §§7–8. Thus, given a modified framework, where models are not of the domain and interpretation function structure, we can substitute the domain in the definitions above with whatever structure we have that is assigned to represent possible worlds and fix terms in that framework. Alternatively, we can settle on just some correspondences to provide relations of representation.

§6. Can any term be fixed? Presumably, debates on logical terms are about which terms should have their meaning fixed across models. Proponents of the demarcational project contend that only logical terms should be fixed, if an apt modeling of logical consequence is what we are after. This would make it seem that for any term with an associated intension, we could either fix it or not, and we should fix it only if the fixed term satisfies the proposed criterion for logical terms. But the semantic framework we are using, of standard extensional logic, imposes considerable restrictions on what *can* be fixed. First, there are the standard examples of intensional operators, such as *necessarily*, *it ought to be that* and *it is believed that*. Their meaning is not a function of just the population of the possible world or situation in which they are evaluated, but in a range of them: that is why each model in semantic systems for these operators includes an elaborate structure of constructs called "possible worlds".

Leaving the intensional operators aside, we are still left with a significant class of terms that cannot be fixed faithfully in the given framework. In §4 I described the relation between models and possible worlds. By *Replacement*, the only requirement of a domain for being used to represent a population of a possible world was that of size. Any one-one and onto correspondence between a domain and the population of a possible world amounts to the representation of the latter by the former. This starting assumption, natural to model-theoretic practice, directly entails a result which is also in agreement with standard practice, by which the terms that are fixed in accordance with some intension are invariant under isomorphisms. This easy result is proved below. Invariance under isomorphisms is one of the most widely accepted criteria for logical terms, and here we see how the criterion is inherent to standard model-theoretic thinking. Note that the argument below pertains only to faithful fixing, so only to terms that have an associated intension. This still leaves room for terms fixed by some operation that is not invariant under isomorphisms, as long as they are not meant to capture an intension.

Let f be a one-one and onto correspondence from D to D'. Let us extend f to apply to set-theoretic constructs over D and D' by recursion in the natural way, as before (§4). We say that an operation O is invariant under isomorphism if for any such extended f, $f(O(D)) = O(D')$. Now we claim that any faithfully fixed term denotes an operation that is invariant under isomorphisms.

For simplicity, we shall confine ourselves to monadic predicates. Let P be a monadic predicate in our language, and assume P is fixed faithfully. Let f be a one-one and onto correspondence from D onto D'. Let w be a possible world with a population whose cardinality is equal to that of D and D'. Let g be a one-one and onto correspondence between D and the population of w, and let $h = g \circ f^{-1}$, so that h is a one-one and onto correspondence between D' and the population of w. Let us extend f, g and h, as before, to apply to set-theoretic constructs over their domains in a natural way. Now, since P is fixed, there is an operation O_P such that for any model $M = \langle D, I \rangle$, $I(P) = O_P(D)$. Since P is fixed faithfully, $g(I(P))$ and $h(I'(P))$ are both the extension of P in w, so $g(I(P)) = h(I'(P))$. So $g(I(P)) = g \circ f^{-1}(I'(P))$, and then, applying g^{-1} and f to both sides, we have $f(g^{-1}(g(I(P)))) = f(g^{-1}(g \circ f^{-1}(I'(P))))$, and so we have $f(I(P)) = I'(P)$, i.e., $f(O_P(D)) = O_P(D')$, proving invariance under isomorphisms.



Let me reiterate the basic assumptions in play, to make the course of the argument maximally explicit. We have assumed that: a) In standard model-theoretic semantics, the interpretation of a fixed term is a function of the domain in each model. b) Models represent possible worlds, and (due to *Replacement*) on each one-one and onto correspondence between a domain of a model and a population of a possible world, there is a relation of representation between individuals and sets built over the respective domain and population. And building on the first two assumptions, we have assumed that: c) A *faithfully* fixed term should be interpreted by an operation that is in full accord with the associated intension. These assumptions lead, through the proof above, to the inevitable conclusion that a term that is faithfully fixed will be invariant under isomorphisms. Each of the assumptions coheres with model theory (as practiced in mathematical logic) and is philosophically sound in the context of this discussion. To be sure, any of these assumptions can be altered or discarded (examples will be given in the next section). But one should provide an explicit working alternative (perhaps giving up standard model-theoretic semantics) if one is to dispute the conclusion of this argument.

For illustration, let *Red* be a predicate in our language associated with the usual (or perhaps a sharpened) meaning from its phonetic correlate in natural language. *Red*, on most accounts of logicality, is regarded as a paradigmatic example of a nonlogical term. But still, even if we were to consider fixing *Red* as a candidate logical term, we would run into trouble. Assuming that there is a possible world w containing both red and nonred objects, there is no consistent way of assigning *Red* an appropriate extension in models with domain D equinumerous with the population of w. Let $d \in D$. There is a one-one and onto correspondence between D and the population of w such that d is mapped to a red object, and one in which it is mapped to a nonred object. We cannot fix *Red* to be a function of the domain without violating its intended meaning. The way we set things up so far, *Red* is simply *precluded* from being a logical term.

The problem with a term such as *Red* is that its meaning differentiates between objects co-inhabiting possible worlds, and in general, it requires distinctions the model-theoretic apparatus, on the given assumptions, cannot make. The same problem does not arise for standard logical terms. The intension of the universal quantifier, understood as "all objects", can be captured by an isomorphism-invariant term. The interpretation of '$\forall$' as a function of the domain doesn't lose any of its intended meaning—it does not erase any of the distinctions '$\forall$', on its intended meaning, could make.

If we stick to the assumptions we have made so far, pertaining to the technical apparatus of models and to the relation between models and possible worlds, we obtain a precondition on logical terms, stemming from our technical and conceptual framework: some terms should not be considered logical or should not be fixed simply because they cannot be fixed. To be sure, we can modify some of the assumptions, as will be demonstrated in the next section. Yet, the difference between *Red* and a standard quantifier such as '$\forall$' remains: fixing the former faithfully requires altering our assumptions while the latter imposes nothing of the sort.

A common way of speaking suggests that if *Red* is to be fixed, then the extension of *Red* in each model should simply be the red things in that model. Indeed, when models are explained in introductory logic classes, they are often presented as having concrete elements. In the current setting, the members of model domains are mathematical objects that represent possibly red things but are not themselves red—this might then seem as an unnecessary divergence from common presentations of model theory, and perhaps a source of the problem I have set out to deal with. But this thought is mistaken. Including concrete entities as elements of models will not solve the issue at hand, and will moreover bring further complications.

First, we note that merely considering concrete objects as they are in the actual world will violate *Necessity*: if *Red* is to be fixed, then presumably, even though my pen is not actually red, it is still red in some possible world, and this should be captured by some model. Thus, a natural move would be to consider domains to be populations of possible worlds. However, in the current extensional setting, the extension of *Red* cannot be determined by just the set of inhabitants of a possible world. Let w be a possible world in which my pen is red. Presumably, we can consider a possible world w' that has exactly the same population as w, but in which my pen isn't red. Now in order for *Red* to be fixed by the technical definition of §5, its extension in a model ought to be a function of the domain. This would violate the intended meaning of *Red*, as we are forced to assign to it the same extension in all models with the same domain as w, regardless of which world we perceive the domain to come from.

The point is that domains are simply sets. My pen cannot be a member of one set as a red pen and of another as a blue pen. Including such information requires more structure than the mere set of objects in a possible world. This information is sometimes thought of as coming from the interpretation function. But recall our definition of fixity: a fixed term's extension in ordinary models must be a function of just the domain. In standard model theory, we are able to fix the standard logical terms through their defining semantic clauses employing the domain as a parameter. I see no way of formulating an adequate semantic clause for *Red*, given the assumptions of *Necessity* and *Replacement*.

Thus, if we were to consider populations of possible worlds themselves to be model-domains, we would need a way of signalling which possible world they come from, and we would need to consider a more intricate conception of models. Indeed, in the next section I propose a way of adding structure to models so that *Red* can be fixed. Nevertheless, I hold on to the approach to models that takes them to be purely mathematical for reasons that were mentioned in §4. It should be clear that resisting this approach and simply referring to the "red things" in the domain, despite the intuitive appeal, does not provide a genuine shortcut for solving our problem: what serves a heuristic purpose in a lax context should be set aside in work aimed to explicate logicality. Yet loosening the restrictions we started out with in an explicit and rigorous way will allow fixing a wider class of terms, as we shall see next.

§7. Adding structure. A term such as *Red* cannot be fixed faithfully with the standard apparatus on our assumptions. However, models can be modified so that in addition to a domain and an interpretation function there is another function which assigns extensions to terms and is designed to represent an intended meaning. So in addition to a mere set representing the population of a possible world, we have some structure imposed on it.

Upon observation of what it takes to fix a term such as Red, we might then conclude that Red should not be fixed, considering the complicated structure and metaphysical assumptions that would be required. Thus, this section should not be seen as recommending the addition of structure. Indeed, I shall later return to the issue of logicality, and propose a graded notion of logicality. Some terms require more structure than others in order to be fixed faithfully, and for this reason can be viewed as less logical—and should not be considered as the fixed terms.

I adapt a framework proposed by Christopher Menzel (1990). The solution is thus to add more structure to the models, so that even before the interpretation function comes in, models with the same domains can still be distinguished, and they will represent different possible worlds. The idea, modified to the current setting, is that each model will be a triple $\langle D, P, I\rangle$ such that D and I are the usual domain and interpretation function, and the added element P is a function from terms to extensions. The first two elements in a model represent a possible world, including a population of objects in that world, represented by D, and the properties and relations between the objects, represented by P. P, the "intended interpretation function", just like the interpretation function I, assigns extensions to terms. But unlike I, P makes no distinction between logical and nonlogical terms. For each term t, $P(t)$ is the extension of t under t's intended meaning. That is, $P(t)$ is the extension of t in the possible world represented by the model, given the meaning t bears. The interpretation function I, too, assigns an extension to t. We shall soon define fixity here, but the idea is that if t is to be fixed in the semantics, I should assign $P(t)$ to t in all models. Otherwise, if t is not fixed, I can assign to t any extension that fits its semantic category (cf. Menzel (1990, pp. 360f)).[10]

With this addition of structure, we can capture more than mere cardinality differences between possible worlds. "[N]ote", Menzel tells us, "that things could have been different in one of two ways: there could have been more or fewer individuals, or the individuals there are could have had different properties and stood in different relations to one another. Distinct plain vanilla models can capture the first sort of difference well enough (simply in virtue of having different domains), but not the second" (Menzel, 1990, p. 359). Going back to the example in the previous section: on the usual assumptions, distinct models cannot capture the meaning of a predicate such as Red. Given *Replacement* in its original form, they can only capture intensions that are simple enough to be represented by isomorphism-invariant terms. And indeed, the latter are terms that have to do with cardinality, a topic I shall elaborate on in §8.

The proposal at hand is to redefine models to have more structure that will be able to represent the particular identities of objects through the differing properties and relations they have in possible worlds. *Replacement* would be revised accordingly, as now we can use models to represent much more than the bare populations of possible worlds. The

[10] Menzel discusses modal logic, and the ability of clusters of models (as in possible worlds semantics) to characterize modal reality—that is, whether we can take a bunch of standard models of extensional logic to represent possible worlds. Menzel is specifically concerned with separating out the change in extension of a predicate due to change in the world and one that is due to change in meaning. We are here concerned with fixing the meaning of terms while still maintaining the variation in extension due to change in possible worlds. Thus, despite the different framework (intensional logic rather than extensional logic), Menzel's discussion is highly relevant to the issue at hand.

pair $\langle D, P\rangle$ is now the component of a model that represents a possible world, taking into account not just its population, but the intended interpretation of terms over that population:

> *Replacement*$^+$: A domain D and intended interpretation function P represent a possible world w if and only if there is a one-one and onto correspondence between members of D and objects in w such that for each term t, the correspondence takes $P(t)$ to t's extension in w according to its associated intension. Each such correspondence constitutes a relation of representation, where each member of D represents an object in w.

Menzel's solution requires modifying model theory: standard models on their own do not suffice to determine a possible world, and therefore another function is added as a component in models. Therefore, we need to change the definition presented in §5: A term t is *fixed* if its interpretation is a function of the domain D and the added function P such that $I(t) = P(t)$.[11] A term associated with an intension is *fixed faithfully* if it is fixed, and its extension via I in each model represents the value of the intension in each possible world represented by the domain of the model D together with P.

Getting a term to be fixed in this framework is not difficult. For a term such as Red, one simply needs to ensure that $I(Red) = P(Red)$ in all models. How do we fix a term *faithfully*? P must then adequately represent the metaphysical profile of the term—in the case of Red, the range of possibilities for red things should be adequately represented.

It becomes obvious that a logic with a term such as Red fixed faithfully involves a fair amount of metaphysics. For instance, the logical status of $\varphi := \neg\forall x Red(x)$, which is neither logically true nor logically false when Red is not fixed, becomes unclear when Red is fixed (assuming '$\forall$' and '$\neg$' are fixed). Presumably, φ would not be logically false when Red is fixed: there is a possible world where not everything is red. However, it is more difficult to tell whether φ is logically *true*: is there a possible world where everything is red? Answering this involves more metaphysics than the present author is informed about.

Another option is to bypass this revision to model theory, while still allowing to fix a term such as Red. We have assumed so far, via *Replacement*, that elements of model domains can represent any objects in possible worlds. Instead of adding an intended interpretation function, perhaps we can simply give up the liberalness of *Replacement*, and lay some more burden on the relation between elements of domains and the objects they represent. A domain will no longer be used to represent any possible world having the same cardinality. One could, for instance, use ordinals to code pairs, the first element of which is an object, and the second a list of properties and relations holding of them. Thus, each object that appears in different possible worlds will be represented by different ordinals, according to the properties it has and the relations it bears to other objects in each of the possible worlds. When we wish to fix a predicate such as Red, we assign to it the operation that given a domain D, it returns the set of ordinals in D coding a pair whose second element includes the property of being red.[12]

It is not clear that such a coding function is feasible. If we were to use ordinals to code all properties and relations whatsoever, including relations of objects to the ordinals

[11] Note that taking these more elaborate structures into account, every fixed term is invariant under isomorphisms between these structures.

[12] Here, isomorphism-invariance is patently lost: while a term such as Red may be fixed, it will remain isomorphism-variant as it is under the standard assumptions.

themselves, we would soon encounter versions of Russell's paradox. We might thus limit the properties and relations that we look at, and limit ourselves only to those expressible in a given restricted language. Still, it should be clear that while the coding solution to the problem of fixing terms keeps models in their standard form, the way models represent possible worlds becomes exceedingly complicated.

In both solutions, the extension of a fixed term is a function of constituents of the model representing a possible world. In the Menzel solution, the extension of a fixed term is a function of $\langle D, P\rangle$. In the coding solution, if feasible, the extension of a fixed term is a function of D alone. In this way we bring *Red* to the same field as '$\forall$' and '$\exists$'.

In both solutions we added some structure to our model-theoretic apparatus: either by complicating models or by complicating their relation to possible worlds. In both cases, while *Red* can in principle be fixed, it is unclear we can obtain a workable semantics. Surely, there might be better frameworks to handle these issues. However, whichever framework we use, if our language has the capacity to express '$\neg\forall x\, Red(x)$', metaphysical complications are imminent. This, I think, is a good reason to rule out fixing *Red* as a logical term. This is the strength of the blended approach (introduced in §4): we avoid metaphysical complications by keeping some terms *un*fixed.

We might extract here a principle regarding logical terms: a term is logical if the machinery required for fixing it is metaphysically acceptable and easy to handle. This principle is in line with approaches by which metaphysical complications should largely be avoided in logic (see Sagi (2014) for examples of such approaches).

The proposed principle of metaphysical modesty does not provide us with a strict criterion for logicality, as it is unclear where to draw the line of metaphysical acceptability. Nonetheless, this suggests a real distinction between terms with respect to their logicality; rather than a categorical account, a graded account emerges. The less structure that is required to faithfully fix a term, the more logical it is. There is an order for logicality, though it may be partial and even vague in some cases. In the next section I propose an absolute measure for the logicality of terms that are invariant under isomorphisms. Those are terms that lend themselves more easily to mathematical treatment. I leave the nature of the order of logicality for nonisomorphism-invariant terms as an open question.[13] To close this section I would rather point out that given some extra structure in our model theory, any term could in principle be fixed. When we allow that extra structure, the obstacle to fixing some terms is our shaky handle on the metaphysics involved.

§8. Logicality, meaning and Löwenheim numbers. We have seen that invariance under isomorphisms is a necessary condition for faithfully fixed terms in first order logic with standard model-theoretic semantics, given our stated assumptions. Isomorphism-invariance thus serves as a distinguished dividing line, separating what can be represented in standard models from what cannot. In the previous section I dealt with one side of the line, the nonisomorphism-invariant terms, and considered modifications to standard model theory that would enable fixing them faithfully. We have seen that adding some structure to models allows them to represent meanings that could not be captured before. The widely accepted criterion of invariance under isomorphisms for logicality receives here support from a new direction, that of the the representation of meaning by models. Terms that

[13] As a first pass, note that necessary properties and necessarily existing immutable objects seem to require of models less structure than contingent properties and objects.

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

can be fixed by isomorphism-invariant operations in standard models are arguably less metaphysically involved than those which cannot.

By this stage, we are able to compare some terms to others and assert that some fare better than others from a logical perspective. It seems to me that in order to conclude that isomorphism-invariant terms are logical tout court, we shall have to assign ordinary models a privileged status, and rule out more stringent criteria. Our results, however, do not warrant such a conclusion, but rather direct us to a graded account of logicality. We may thus consider model theories with *less* structure than the standard one, in which not all isomorphism-invariant terms can be faithfully fixed. One might ask for a *measure* of logicality of terms, following the comparisons we have made. I shall address this request shortly. However, one qualification is in place. We have been considering here one perspective on logicality, one aspect of meaningful terms by which their logicality can be compared. There are, however, various considerations for choosing fixed terms for a system, i.e., for demarcating the logical terms. It is not at all clear that intuitions regarding logicality lead to one coherent notion. As put by van Benthem, "In fact, the standard examples have so many nice properties together, of quite diverse sorts, that it may not even be reasonable to think of their conjunction as defining one single 'natural kind'." (van Benthem, 1989). I suggest that ultimately, the choice of fixed terms is pragmatic, but one of the major considerations at play is the very real distinctions presented here.[14]

Now, I have proposed the principle: the less structure required for fixing a term faithfully, the more logical it is. This theoretical principle can lend itself to further sharpening. I shall propose here one way of sharpening this principle. No doubt, there may be others. The main virtue of the present proposal is that it captures the philosophical idea in a way that is congruent with contemporary model theory.

As we shall be using contemporary model-theoretic machinery, we shall be looking specifically at one side of the dividing line: that of the isomorphism-invariant terms. Contemporary model theory is strictly situated on this side of the line.[15] The study of generalized logics attempts to characterize various logics extending the language of standard first order logic with various isomorphism-invariant terms. Now, according to a common characterization of isomorphism-invariant terms, those terms are concerned only with *cardinality* (Tarski, 1986; Shapiro, 1998). Such terms are indifferent to the identities of particular objects, but are sensitive to the sizes of extensions. In what follows, I build on this idea and propose that among the isomorphism-invariant terms (where, loosely speaking, only cardinality is a factor), a term is more logical the less it distinguishes between different cardinalities.

One way of measuring the extent to which a term distinguishes between cardinalities is by Löwenheim numbers. A logic has a Löwenheim number κ if every sentence that is satisfied by a model of cardinality greater than κ is also satisfied by a model of cardinality at most κ: the logic does not distinguish between models of cardinality from a certain point onwards. If the only structure allowed in determining the meaning of isomorphism-invariant terms is cardinality, in a logic with Löwenheim number κ all the structure that is

[14] In this I follow Sher, who explains: "My considered view is this: If the question is "Which logical constants should we include in the logical system we are working with?", then this question has multiple dimensions, and at different times we should make different decisions, depending on what dimensions are most important to us and what our goals are at those times" (Sher, 2013, p. 182). However, while Sher offers a strict criterion for logicality on one of the said dimensions, I propose a graded account.

[15] Accordingly, I shall assume the axioms of ZFC throughout.

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

allowed is cardinality-up-to-κ. I shall mention alternative notions related to the indifference of logics to size, but will argue that in an important sense, Löwenheim numbers have an advantage in explicating our theoretical principle.[16]

We have mentioned the Tarskian dictum that isomorphism-invariant terms are concerned only with size. We can explicate this idea in the following manner. Let us consider a slight change to standard model theory, for the sake of evaluating isomorphism-invariant terms. We make the class of models "thinner": in the new version, all models of the same cardinality have the same domain—so there is only one domain of each cardinality. In fact, since cardinal numbers are sets, we can simply use cardinal numbers as the domains of all models (but this particular choice is arbitrary: any choice of a set for each cardinality will give the same results). For standard languages, where the fixed terms are isomorphism-invariant, this revision does not affect validity: the arguments that are valid in the ordinary version are exactly those that are valid in the thin one. So the thinner model theory is simply a more economic version of ordinary model theory, without what may be considered as an excess in models. This is a result of the following two facts about logics where the logical terms are isomorphism-invariant and the full class of models is employed:

Let L be a language where the logical terms are isomorphism-invariant.

FACT 8.1. Let D be a set and $M' = \langle D', I' \rangle$ a model for L such that D and D' have the same cardinality. Then there is a model $M = \langle D, I \rangle$ for L that is isomorphic to M'.

FACT 8.2. Every two isomorphic models for L are elementary equivalent (satisfy the same sentences): if $M \cong M'$ then $M \equiv M$.

Now, as we changed the class of models, we must attend to the relation between the thin class of models to possible worlds. Here, as before, in order to represent a possible world w we use a domain of the same size as the population of w.

Considering the earlier discussion on what it takes to fix a term, we might worry whether the isomorphism-invariant terms can be fixed in the thin model theory in a manner faithful to their meaning—whether their meaning is not somehow violated when some models are lost. The claim before was that standard model theory is not capable of fixing *Red* faithfully, but the standard logical terms could nonetheless be fixed. The question now is whether the isomorphism-invariant terms can be fixed faithfully in the thin, revised model theory. I claim that if they were faithfully fixed in the standard system, they can be fixed faithfully in the revised one, and that their meaning is not violated. Firstly, in the thin model theory, we cut down the models, but we still have enough of them to represent all possible worlds, so there is no worry that some aspect of the meaning is lost. Furthermore, these terms can still be faithfully fixed precisely because they "behave" the same in all models of the same size in the standard system.

This exercise shows that the difference between distinct equinumerous domains is immaterial for the sake of capturing the meaning of isomorphism-invariant terms, and can be treated as excessive structure. We have learned that not only does the thin model theory give us the same validities as the standard one, it is faithful to the meanings of the logical vocabulary we standardly use. And this vocabulary is very economical in the amount of structure it requires of the model theory—whereas other terms, such as *Red*,

[16] I am indebted to Menachem Magidor, who suggested looking into Löwenheim and Hanf numbers.

require much more structure to appropriately model their behavior in possible worlds. Indeed, Sher, as other proponents of invariance criteria, has claimed to have captured the *generality* of logical terms (Sher, 1996; Bonnay, 2008). We see here another aspect of this generality: isomorphism-invariant terms do not require commitment to a specific metaphysically elaborate set-up. They can be captured by a simple, mathematically and metaphysically undemanding model theory.

The literature on logicality contains some dissatisfaction with the criterion of isomorphism-invariance, mostly tied with the complaint that it overgenerates. As an alternative, stricter invariance criteria have been proposed (Feferman, 1999; Bonnay, 2008). Such invariance criteria provide a way of distinguishing between isomorphism-invariant terms. Here, I shall take a different route that follows directly from the present considerations. We shall, however, see how well alternative invariance criteria fare with respect to the present proposal.

Now, given our thin class of models, further reductions will provide further distinctions among isomorphism-invariant terms. We proceed with some definitions. One fundamental notion for present purposes is that of *spectrum*, which tells us for a given formula in a given logic the range of cardinalities of models that satisfy it.

DEFINITION 8.3 (Spectrum). *For a formula φ in a logic L,*

$$sp(\varphi) = \{|M| : M \models_L \varphi\},$$

where $|M|$ is the cardinality of the domain of M.

The notion of spectrum is directly connected to that of a Löwenheim number. Let us define Löwenheim and Hanf numbers.

DEFINITION 8.4. *Let L be a logic.*

1. *The Löwenheim number of L, $\ell(L)$, is the least cardinal μ such that any satisfiable sentence in L has a model of cardinality less or equal to μ if such exists. Otherwise, $\ell(L)=\infty$.*
2. *The Hanf number of L, h(L), is the least cardinal μ such that for any sentence in L, if it is satisfiable by a model of cardinality greater or equal to μ, then it is satisfiable by arbitrarily large models.*

Considering specifically Löwenheim numbers and Hanf numbers, we observe that they can be defined by the notion of a spectrum:

FACT 8.5. Let L be a logic.[17]

1. If the Löwenheim number of L exists, then
   ℓ(L)= $sup\{min\ C :$ C is a spectrum of a sentence in L$\}$.
2. If the Hanf number of L exists, then
   h(L)=$sup\{sup\ C :$ C is spectrum of a sentence in L that is bounded from above$\}$.

[17] Indeed, some authors use the following statements as definitions of Löwenheim and Hanf numbers, e.g., (Väänänen, 1985, pp. 637, 639). The statement on the Löwenheim number of a logic readily follows from our definition. The statement on the Hanf number of a logic follows from our definition, and the claim that if $sup\{sup\ C :$ C is spectrum of a sentence in L that is bounded from above$\}$ exists, it is a limit cardinal that does not belong to any bounded spectrum of a sentence in the logic (Ebbinghaus, 1985, pp. 64f).

Both Löwenheim and Hanf numbers tell us something about the extent to which a logic discriminates between different cardinalities. No spectrum of a sentence can lie strictly above the Löwenheim number of the logic. And the Hanf number of a logic is a cardinal number which whenever it belongs to a spectrum of a sentence in the logic, so do arbitrarily large cardinalities above it.

We can immediately observe that by the downwards Löwenheim-Skolem theorem, standard first order logic has Löwenheim number $\aleph_0$, and by the upward Löwenheim-Skolem theorem it has Hanf number $\aleph_0$. We shall soon give further examples of Löwenheim numbers. We note at this point that Hanf numbers prove to be much more difficult to compute, and in many cases their existence has been proven to depend on strong large cardinal axioms. For this reason and for the reason we give next, we shall henceforth focus our attention on Löwenheim numbers.

Löwenheim numbers are of particular interest to us, as they allow cutting down the class of models for given logics. That is, for a logic L, if $\ell(\mathrm{L})=\kappa$, then dismissing all models of cardinality greater than κ will not make a difference to the validities of the logic. Combined with our earlier observation, this means that it is sufficient to take one domain of each cardinality less or equal to κ to account for all countermodels (this is easy to see when considering the notion of spectrum: no spectrum of a sentence in the logic distinguishes between cardinalities above κ). We may venture to say that the terms in the logic L in such a case, if fixed faithfully, require no more structure than that which is given by the set of cardinalities less or equal to κ. For example, the terms in the logic $\mathcal{L}(Q_\alpha)$ (to be defined below) do not distinguish between cardinalities greater than $\aleph_\alpha$ and thus require no more structure than is given by the set of cardinalities less or equal to $\aleph_\alpha$. Hanf numbers do not lend us a way of reducing the class of models, and while they are of interest, they are not as useful as Löwenheim numbers for our purpose.

As before, one might worry that if we cut down the class of models, even if validity has not been affected, some content will be lost. Here the worry is more substantial. Assume that we have a logic with Löwenheim number κ, and so we exclude from the class of models all models of size $> \kappa$. As we take possible worlds to be represented by models of the same cardinality, we presumably have now lost representation of all possible worlds of size $> \kappa$. By the first thinning to one domain of each cardinality, we have reached the limit allowable by *Replacement* and the assumption that there are possible worlds of all cardinalities. Barring a metaphysical argument to the effect that there are no possible worlds of size $> \kappa$, our notion of representation needs to be altered. So to account for large possible worlds, we shall lean on the expressive power of the language. The fact that a sentence satisfied by a model is also satisfied by a model of size $\leq \kappa$ tells us that all the meaning that was captured in the full range of models can still be captured when they are cut down.

In this modified setting, we should again define the fixed and the faithfully fixed terms. In a reduced class of models of size $\leq \kappa$, a term is said to be *fixed*, as before, if its interpretation is a function of the domain. A term is *fixed faithfully* in a class of models of size $\leq \kappa$ if it is fixed, and furthermore, the operation fixing it can be extended to the full range of models so that it is faithfully fixed there (according to the definition in §5), and its associated Löwenheim number when thus extended is $\leq \kappa$. Note that a Löwenheim number can be assigned to a term only on the backdrop of a language, as we shall now see.

Let us look at some more examples of Löwenheim numbers. For a logic L and a quantifier Q, let L(Q) be the logic L with Q added to its language. Let $\mathcal{L}$ denote standard first order logic.

- Let Q_α be the unary monadic[18] quantifier "there are at least $\aleph_\alpha$ many". We have $\ell(\mathcal{L}(Q_\alpha))=\aleph_\alpha$ for each ordinal α (Väänänen, 1985, p. 637).
- Let Q^W be the unary polyadic quantifier over binary relations such that $M \models Q^W xy\varphi(x, y)$ iff $\varphi(x, y)^M$ is a well-order. We have $\ell(\mathcal{L}(Q^W))=\aleph_0$ (*ibid*).
- Let the Härtig quantifier I be the binary monadic quantifier stating equal cardinality of sets: I is a binary monadic quantifier such that $M \models Ix(\varphi x, \psi x)$ iff $|(\varphi x)^M| = |(\psi x)^M|$. $\ell(\mathcal{L}(I))$ is very high, and is independent of ZFC. $\ell(\mathcal{L}(I))$ is a fixed point of the function $\alpha \mapsto \aleph_\alpha$, and further, Magidor and Väänänen showed that it is consistent with ZFC both that $\ell(\mathcal{L}(I))$ is under the first weakly inaccessible cardinal and that it is above the measurable cardinal (Magidor & Väänänen, 2011).
- Let Most be the binary monadic quantifier such that $M \models \mathrm{Most}\, x(\varphi x, \psi x)$ iff $|(\varphi x)^M \setminus (\psi x)^M| < |(\varphi x)^M \cap (\psi x)^M|$. Then we have: $\ell(\mathcal{L}(\mathrm{Most})) = \ell(\mathcal{L}(I))$ (See Jeřábek (2016)).
- Let More be the binary monadic quantifier such that $M \models \mathrm{More}\, x(\varphi x, \psi x)$ iff $|(\varphi x)^M| > |(\psi x)^M|$. Then we have $\ell(\mathcal{L}(\mathrm{More})) = \ell(\mathcal{L}(I))$.[19]

Löwenheim numbers will be the key for the measure for logicality. However, Löwenheim numbers apply to *logics*, not to terms, and it is terms whose logicality we measure here by what it takes to fix them faithfully. We shall thus view Löwenheim numbers as telling us how much structure a term requires in order to be fixed *in the context of a logic*. Having given up full representation by cutting down models, the expressive power of the language gets factored in. I take standard first-order logic to be a reasonable background context, although I shall consider alternatives later on. Thus, restricting ourselves to quantifiers for the sake of simplicity, the logicality of a quantifier Q will be measured by $\ell(\mathcal{L}(Q))$.[20]

We note that for any Q, $\ell(\mathcal{L}(Q))$ exists (the reason being that the sentences of the logic constitute a set; some infinitary logics, to be considered below lack this property). So, for example, for $\alpha \leq \beta$, $\ell(\mathcal{L}(Q_\alpha)) = \aleph_\alpha \leq \aleph_\beta = \ell(\mathcal{L}(Q_\beta))$, and so Q_α is *more logical* than Q_β. Moreover, we see that various quantifiers that are used for comparison of size such as Most, More and I can be shown to have the same, quite high, Löwenheim number.[21] Bringing in metaphysical motivation, we explain that the higher set-theoretic infinite is metaphysically loaded. The lower the cardinalities to which the meaning of a term may be sensitive, the more logical it is.

---

18 A quantifier is *monadic* if its arguments are unary predicates, and is *polyadic* otherwise (if at least one of its arguments is a relation).

19 Note that due to considerations of expressive power, $\ell(\mathcal{L}(\mathrm{More})) \geq \ell(\mathcal{L}(I))$. I.e., I is expressible in $\mathcal{L}(\mathrm{More})$: $Ix(\phi x, \psi x)$ is equivalent to $\neg\,\mathrm{More}\, x(\phi x, \psi x) \wedge \neg\,\mathrm{More}\, x(\psi x, \phi x)$. For the other direction, that $\ell(\mathcal{L}(I)) \geq \ell(\mathcal{L}(\mathrm{More}))$, see Jeřábek (2016).

20 We should note that formal criteria for logicality are invariably affected by choices regarding the formal framework. For example, Casanovas shows how some invariance criteria yield different results if a relational or functional types framework is chosen, and he draws the conclusion that "the logical character of notions and operations from a semantical point of view is a matter of perspective and a matter of degree" (Casanovas, 2007, p. 36). In the present context, we follow common practice by choosing first-order logic as the background logic, without barring well-justified alternatives.

21 Note that Most, More and I are not all interdefinable. Indeed, we have the strict ordering $\mathcal{L}(I) < \mathcal{L}(\mathrm{More})$ and $\mathcal{L}(\mathrm{Most}) < \mathcal{L}(\mathrm{More})$ on the order of expressive power, while $\mathcal{L}(I)$ and $\mathcal{L}(\mathrm{Most})$ are incomperable. See Westerståhl (1989).

We should distinguish between a logic $\mathcal{L}(Q)$ used to measure the logicality of Q and the logic we ultimately use for validity and logical consequence. Based on the measure provided by $\mathcal{L}(Q)$ for various Q's we then choose which of them to fix—and here enter other considerations, depending on the purpose at hand. For instance, the logic $\mathcal{L}(Q_1)$ happens to have some nice properties lacking in $\mathcal{L}(Q_0)$—having, for example, a complete proof system. Add to this that we do not provide a strict criterion that will categorically determine the logical terms. Thus, when choosing a logic, one takes into account the logicality of the terms and other considerations, and weighs them with respect to the purpose at hand. Again, I would like to gesture here towards a pragmatic approach to logical consequence (that I cannot fully develop here). At the same time, the logicality of terms is deemed a nonpragmatic matter which has its own measure.

Now we might wish to evaluate model-theoretic criteria for logicality that have been proposed in the literature in light of our considerations. I shall consider three invariance criteria: invariance under isomorphism (commonly referred to as the *Tarski-Sher criterion*), invariance under potential isomorphism (due to Bonnay), and invariance under homomorphism (due to Feferman). The question regarding all of them will be: is there a bound to the Löwenheim number of $\mathcal{L}(Q)$ given a Q that satisfies the criterion?

Invariance under isomorphism. Recall that invariance under isomorphism was our starting point in this section. Any possible quantifier can be considered here, as long as it regards only cardinalities. Now note that there is no common Löwenheim number to $\mathcal{L}(Q_\alpha)$ when every α is considered. So, there is no "least logical" term in the class we are considering.

Now, it would seem appropriate to match the criterion of isomorphism-invariance with the infinitary logic $L_{\infty\infty}$ (which allows infinite set-sized disjunctions and quantification over infinite set-sized sets of variables), since the terms locally definable in that logic are exactly those that are isomorphism-invariant (McGee, 1996). One might try to measure the logicality of the terms satisfying this criterion by looking at the Löwenheim number of $L_{\infty\infty}$. However, $\ell(L_{\infty\infty}) = \infty$ (note that $L_{\infty\infty}$ has class-size many sentences). The fact that the Löwenheim number of $L_{\infty\infty}$ is "maximal" is consistent with our present starting point being invariance under isomorphisms. Nonetheless, it does not seem that evaluating an invariance criterion should be identified with evaluating the logic in which the terms passing the criterion are definable, if that logic is infinitary. The proposal of setting isomorphism-invariance as a criterion for logicality is merely a proposal that only terms that are isomorphism-invariant should be fixed, not that the right logic is infinitary (although those claims would be consistent, and it seems that McGee supports something of the latter sort (McGee, 1996). Allowing infinitely long sentences makes a substantive difference to the expressive power of a logic. Definability results should thus serve as a tool or seen as a bound, and not seen as providing the right logic according to an invariance criterion. Thus, it appears that the logics advocated by the proponents of invariance under isomorphism are of the form $\mathcal{L}(Q_1, \ldots, Q_n)$ where $Q_1, \ldots, Q_n$ are invariant under isomorphisms. The logicality of each of the quantifiers is measured separately—and indeed, the respective Löwenheim number always exists—but we see that satisfying the Tarski-Sher criterion does not entail a bound on the Löwenheim number (because in general, Löwenheim numbers are boundless).[22]

[22] Here we focus our attention on single terms (taken one by one), and we measure their logicality given the constant background context $\mathcal{L}$. The considerations leading to the logic one chooses at

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press

Invariance under potential isomorphism.[23] Considered in maximal generality, the criterion of invariance under potential isomorphism does not yield better results than that of invariance under isomorphisms. Thus, we have a stricter criterion (not all isomorphism-inavriant terms are invariant under potential isomorphisms), but no real advantage when it comes to distinguishing between infinite cardinalities. As an example, consider for an ordinal α, the *ordinal quantifier* over binary relations $Q_{\cong\alpha}$ which applies to relations if and only if they have order-type α. There is no bound on the Löwenheim number of $\mathcal{L}(Q_{\cong\alpha})$ when all ordinals α are considered. However, we obtain a better result when we restrict ourselves syntactically: Let Q be a monadic quantifier (acting on unary predicates). Then $\ell(\mathcal{L}(Q))= \aleph_0$.[24,25]

the end include the measure of logicality of the terms that get fixed, but also properties of the logic as a whole—which we are able to mention here only in passing.

23 The definition of *potential isomorphism* relies on the one for *partial isomorphism*:

DEFINITION (Partial isomorphism). *Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures, and $f : \mathcal{M} \to \mathcal{M}'$ a function. f is a partial isomorphism between $\mathcal{M}$ and $\mathcal{M}'$ iff there are two substructures $\mathcal{A}$ and $\mathcal{A}'$ of $\mathcal{M}$ and $\mathcal{M}'$ such that f is an isomorphism between $\mathcal{A}$ and $\mathcal{A}'$.*

DEFINITION (Potential isomorphism). *A potential isomorphism I between two structures $\mathcal{M}$ and $\mathcal{M}'$ (notation $I : \mathcal{M} \cong_p \mathcal{M}'$) is a nonempty set of partial isomorphisms such that: for all $f \in I$ and a in the domain of $\mathcal{M}$ (resp. b in the domain of $\mathcal{M}'$), there is a $g \in I$ with $f \subseteq g$ and $a \in dom(g)$ (resp. $b \in rng(g)$).*

See Bonnay (2008, p. 46).

24 Outline of proof: We note that structures of monadic type of arity n can be represented by a partition of the domain into 2^n cells, such that the cardinalities of the cells completely determines their isomorphism type. With potential isomorphisms, it suffices to consider for each cell whether it is finite or infinite, and only if it is finite, its exact cardinality.

That is, let $\mathcal{M} =< M, A_0, \ldots, A_{n-1} >$ be an n-ary monadic structures. If $X \subseteq M$, let $X^0 = X$ and $X^1 = M \backslash X$. If s is a function from $\{0, \ldots, n-1\}$ to $\{0, 1\}$, let $P_s^M = A_0^{s(0)} \cap \cdots \cap A_{n-1}^{s(n-1)}$. $\{P_s^M\}$ is a partition of M, and, up to isomorphism, the number of elements in these partition sets is all there is to say about M (see Westerståhl (1989, p. 25)). Now consider a countable structure $\mathcal{M}' =< M', A'_0, \ldots, A'_{n-1} >$ such that for each s, $P_s^{M'}$ has the same number of elements as P_s^M in case the latter is finite, and $\aleph_0$ many elements otherwise. $\mathcal{M}$ and $\mathcal{M}'$ are potentially isomorphic. This entails that every n-ary monadic structure is potentially isomorphic to a countable structure. One can then prove by induction on the formulas of $\mathcal{L}(Q)$ where Q is invariant under potential isomorphisms that every sentence satisfiable in an infinite model is satisfiable in a countable model, i.e., that $\ell(\mathcal{L}(Q)) = \aleph_0$.

We might mention in this regard that Bonnay provides a series of results in support of invariance under potential isomorphisms as a criterion for logicality. One of the main philosophical motivations Bonnay appeals to is that not all mathematical truths are logical truths: while it is alright for a logic to distinguish between finite cardinalities (so all arithmetical truths are logical truths), "only quantifiers which do not distinguish among infinite cardinalities pass the test [proposed by Bonnay]" (Bonnay, 2008, p. 64). However, these remarks apply only to monadic quantifiers, since, as Bonnay himself acknowledges, ordinal quantifiers do distinguish among infinite cardinalities. Now, if as a primary aim we wish to make sure that infinite cardinalities are not distinguished, we can just set a requirement for the logic directly that its Löwenheim number will be bound by $\aleph_0$ (and similarly, we may require that the Hanf number will be bound by $\aleph_0$ as well). However, Bonnay is interested in finding a similarity relation between models invariance under which will serve as a criterion for logicality. We note that in the context of the present discussion, invariance under various transformations is of no independent interest, and a direct limitation of the Löwenheim number is best suited for the considerations we have here of measuring logicality.

25 Note here that the terms that are invariant under potential isomorphisms are exactly those locally definable in $L_{\infty\omega}$ (the infinitary logic which allows infinite set-sized disjunctions but only

Invariance under homomorphism.[26] Feferman has proposed that the right kind of transformations to look at are *homomorphisms*. However, the criterion he ultimately proposes (in Feferman (1999)) is not simply invariance under homomorphisms, but rather the following: an operation is logical if and only if it is λ-definable from monadic homomorphism-invariant operations. We shall not go into Feferman's reasoning, but we shall note one distinctive feature of both invariance under homomorphisms and Feferman's modified criterion: the identity predicate is ruled as nonlogical. Thus far we have assumed our candidate logics to include identity. By setting standard first order logic as the background context, we may have assumed too much, since by default all the terms of $\mathcal{L}$ will be maximally logical. In particular, the terms of $\mathcal{L}$ are all put in the same basket. Now, Feferman acknowledges that identity "has a "universal," accepted, and stable logic", and he indeed suggests a way to brute-force identity back (Feferman, 1999, p. 44). But here we are interested in a more fine-grained approach, not just in a categorical distinction. And as Feferman's criterion suggests, it is not at all clear that identity should have the same status as the standard quantifiers '$\forall$' and '$\exists$'. The latter do pass Feferman's test. Indeed, taking identity out of first order logic affects its range of spectra, and more specifically:

Let L be a language where the logical terms are definable from FOL$^-$.

FACT 8.6. Let D be an infinite set and $M' = \langle D', I' \rangle$ a model for L. Then there is a model $M = \langle D, I \rangle$ for L such that $M \equiv M'$.[27]

We note that the terms satisfying Feferman's criterion are exactly those definable in FOL$^-$. The class of models can thus be thinned so that all the models share the same (infinite) domain, and the logical terms are constant across all models in the thin class. Again, I claim that terms denoting operations satisfying Feferman's criterion do not lose any of their intended meaning in the thinned class of models: any intension that can be represented by an operation definable in FOL$^-$ is insensitive to sizes possible worlds.[28]

---

quantification over finite sets of variables) (See Bonnay (2008, p. 62) and Barwise (1985, p. 58)). Now, $\ell(L_{\infty\omega}) = \infty$ (again, note that this logic has class-size many sentences), but as we noted before, this result is not the right one to use when evaluating the criterion of invariance under potential isomorphisms.

26 DEFINITION (Homomorphism). Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures of the same signature σ, a homomorphism from $\mathcal{M}$ onto $\mathcal{M}'$ is a surjective function $F : \mathcal{M} \to \mathcal{M}'$ such that:

- For any constant a in σ, $f(a^{\mathcal{M}}) = a^{\mathcal{M}'}$.
- For any function symbol g of arity n in σ, for any n-uplet $\vec{a}$ of elements in the domain of $\mathcal{M}$, $f(g^{\mathcal{M}}(\vec{a})) = g^{\mathcal{M}'}(f(\vec{a}))$.
- For any relation symbol R of arity $n > 0$ in σ, for any n-uplet $\vec{a}$ of elements of $\mathcal{M}$, $\vec{a} \in R^{\mathcal{M}}$ iff $f(\vec{a}) \in R^{\mathcal{M}'}$.
- For any predicate symbol p of arity 0 in σ, $p^{\mathcal{M}} = p^{\mathcal{M}'}$.

For uniformity, we employed Bonnay's formulation for what he refers to as *strong homomorphism*, see (Bonnay, 2008, pp. 41f).

27 To account for models with infinite domains, we make use of the Löwenheim-Skolem theorem for standard first order logic. To account for finite models, we construct a model $M = \langle D, I \rangle$ together with a homomorphism from D onto D' and prove elementary equivalence by induction.

28 This is not to say that FOL$^-$ as a whole is insensitive to size. For each $\kappa \leq \aleph_0$ there is a theory in FOL$^-$ that is satisfied only by models of size at least κ. Indeed, that is why an infinite model was required in the fact above.

The main difference between FOL$^-$ and first order logic with identity, from our perspective, is that the former does not distinguish between finite cardinalities. Until now we focused on infinite cardinalities, as they seem to carry with them more metaphysical weight. In principle, however, one might wish logic to measure the logicality of terms also with respect to the dependence of their meaning on finite cardinalities. Here I shall take the stand that while FOL$^-$ lies at the minimal end for our measure, since identity is a basic ingredient in logical practice, $\mathcal{L}$ serves as our regular background context.

To sum up this section, we have seen a measure for the logicality of isomorphism-invariant terms. The main focus was on quantifiers. I assumed the backdrop of standard first order logic, measuring the logicality of a quantifier by the Löwenheim number of $\mathcal{L}(Q)$: the lower it is, the more logical is Q—the less Q's meaning is involved with distinctions among infinite cardinalities. The minimal value in this setting is $\aleph_0$, the Löwenheim number of $\mathcal{L}$ itself. To distinguish between '=' and the quantifiers '$\forall$' and '$\exists$', one can look at the spectra of FOL$^-$, observing that this logic avoids distinctions between finite cardinalities as well. The focus on logics at least as strong as $\mathcal{L}$ is due to the fact that when dealing with isomorphism-invariant terms, we have entered a mathematical realm, where $\mathcal{L}$ is assumed in the background.

Looking back at the examples, we learn that the quantifier Q_0—"there are infinitely many"—is maximally logical ($\ell(\mathcal{L}(Q_0)) = \aleph_0$); less logical we have Q_n for other natural n ($\ell(\mathcal{L}(Q_n)) = \aleph_n$); and well beneath those in the level of logicality we have quantifiers as Most, More and the Härtrig quantifier I. One might find it curious that quantifiers that are in pervasive use in natural language and appear quite harmless turn out to be less logical than "there are $\aleph_{2^{64}}$ many", stating a size that is beyond reach in any ordinary setting. However, we note that while in ordinary settings the higher infinite is not considered, Most, More and I when explicated formally, have a meaning that is highly entangled with set theory. We learn that the meaning of Most is significantly more complex than that of the ordinary quantifiers and any Q_n cardinality quantifier—capturing it requires more elaborate structure—and it is thus less logical. Nonetheless, being invariant under isomorphisms, Most is still more logical than any term which is not, such as Red.

§9. Conclusion. The aim of this paper was to incorporate meaning into the philosophical discussion on logicality in a rigorous way. The functional role of logical terms requires them to be fixed. I have offered an explication according to which, in the framework of model-theoretic semantics, what gets fixed is the intension of the logical term. Some terms have more complex intensions than others, and require more elaborate structures in order to be fixed: they are thus less logical.

Observing that on some natural assumptions, only isomorphism-invariant terms can be faithfully fixed in standard model theory, the condition of invariance under isomorphisms provides a natural dividing line. The terms that are invariant under isomorphisms can be handled by mathematical machinery, and so on one side of the dividing line we find that we can give an absolute measure for logicality using Löwenheim numbers. I leave the issue of comparing various nonisomorphism-invariant terms to future work.

Logical terms ultimately serve for a definition of logical consequence and logical truth. What are the implications of this graded account for logicality on these notions? The connection is the following: the more logical the terms that you fix in a system, the less the resulting consequence relation is metaphysically involved. In this way we address Warmbrōd's initial question cited in §2: why should logical terms get fixed? Indeed, it can be argued that a demand for extra structure on the part of the model representing possible worlds may be an indication of metaphysical complications, and it has adverse implications for logic. A good account of logical consequence and logical truth is one that is free of excessive structure coming from complication in metaphysics or meaning.

What we obtain is neither a strict criterion for logicality, nor an ultimate definition of logical consequence, or what might be "the correct logic". I have gestured towards a pragmatic perspective on the matter. A logical system can be chosen for different purposes according to various considerations, in a balance between expressive power and metaphysical modesty. However, the logicality of terms—how metaphysically modest they are—is not a pragmatic issue, and is determined by the amenability of their meaning to formal rendering.

The account presented here is wholly model-theoretic. Surely, there are other frameworks explicating logical consequence (proof theory, among others). In an explication, as Carnap has taught us, there is a certain freedom, given the aims of providing a fruitful and precise account that captures the gist of the phenomenon in a simple way (Carnap, 1962). The account offered here does not only include a fruitful and precise formal framework, but also explains exactly how the formalism relates to the philosophical notions. If other frameworks can accomplish as much, it would indeed be interesting to see whether and how their results may differ.

§10. Acknowledgements. I am grateful to the following people who have commented on earlier versions of this paper: Tim Button, Salvatore Florio, Peter Koellner, Ran Lanzet, Hannes Leitgeb, Christopher Menzel, Stewart Shapiro, Jack Woods, and an anonymous referee for this journal. I've greatly benefitted from conversations with Denis Bonnay, Menachem Magidor, and Kobi Peterzil. Research leading to this paper was supported by the Alexander von Humboldt Foundation.

BIBLIOGRAPHY

Barwise, J. (1985). Model-theoretic logics: Background and aims. In Barwise, J. and Feferman, S., editors. *Model-Theoretic Logics*. New York: Springer, pp. 3–24.

Bolzano, B. (1929). *Wissenschaftslehre*. Leipzig: Felix Meiner.

Bonnay, D. (2008). Logicality and invariance. *The Bulletin of Symbolic Logic*, **14**(1), 29–68.

Carnap, R. (1962). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Casanovas, E. (2007). Logical operations and invariance. *Journal of Philosophical Logic*, **36**(1), 33–60.

Dutilh Novaes, C. (2012). Reassessing logical hylomorphism and the demarcation of logical constants. *Synthese*, **185**(3), 387–410.

Ebbinghaus, H. (1985). Chapter II: Extended logics: The general framework. In Barwise, J. and Feferman, S., editors. *Model-Theoretic Logics*. New York: Springer, pp. 25–76.

Etchemendy, J. (1990). *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.

Etchemendy, J. (2006). Reflections on consequence. In Patterson, D., editor. *New Essays on Tarski and Philosophy*. Oxford: Oxford University Press, pp. 263–299.

Feferman, S. (1999). Logic, logics and logicism. *Notre Dame Journal of Formal Logic*, **40**(1), 31–55.

Gómez-Torrente, M. (1996). Tarski on logical consequence. *Notre Dame Journal of Formal Logic*, **37**(1), 125–151.
Gómez-Torrente, M. (2002). The problem of logical constants. *The Bulletin of Symbolic Logic*, **8**(1), 1–37.
Hanson, W. H. (1997). The concept of logical consequence. *The Philosophical Review*, **106**(3), 365–409.
Jeřábek, E. (2016). Lowenheim numbers for ordinary language quantifiers. MathOverflow. Available at: http://mathoverflow.net/q/249284 (version: 2016-09-07).
Kuhn, S. (1981). Logical expressions, constants, and operator logic. *The Journal of Philosophy*, **78**(9), 487–499.
Magidor, M. & Väänänen, J. (2011). On Löwenheim–Skolem–Tarski numbers for extensions of first order logic. *Journal of Mathematical Logic*, **11**(01), 87–113.
McCarthy, T. (1981). The idea of a logical constant. *The Journal of Philosophy*, **78**(9), 499–523.
McGee, V. (1996). Logical operations. *Journal of Philosophical Logic*, **25**, 567–580.
Menzel, C. (1990). Actualism, ontological committment, and possible worlds semantics. *Synthese*, **85**(3), 355–389.
Mostowski, A. (1957). On a generalization of quantifiers. *Funamenta Mathematicae*, **44**(1), 12–36.
Peacocke, C. (1976). What is a logical constant? *The Journal of Philosophy*, **73**(9), 221–240.
Read, S. (1994). Formal and material consequence. *Journal of Philosophical Logic*, **23**(3), 247–265.
Sagi, G. (2014). Models and logical consequence. *Journal of Philosophical Logic*, **43**(5), 943–964.
Shapiro, S. (1998). Logical consequence: Models and modality. In Schirn, M., editor. *The Philosophy of Mathematics Today*. Oxford: Oxford Univerity Press, pp. 131–156.
Sher, G. (1991). *The Bounds of Logic: A Generalized Viewpoint*. Cambridge, MA: MIT Press.
Sher, G. (1996). Did Tarski commit ‘Tarski’s fallacy’? *The Journal of Symbolic Logic*, **61**(2), 653–686.
Sher, G. (2013). The foundational problem of logic. *Bulletin of Symbolic Logic*, **19**(2), 145–198.
Tarski, A. (1936). On the concept of logical consequence. In Corcoran, J., editor. *Logic, Semantics, Metamathematics*. Indianapolis: Hackett (1983), pp. 409–420.
Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, **7**, 143–154.
Väänänen, J. (1985). Chapter XVII: Set-theoretic definability of logics. In Barwise, J. and Feferman, S., editors. *Model-Theoretic Logics*. New York: Springer, pp. 599–643.
van Benthem, J. (1989). Logical constants across varying types. *Notre Dame Journal of Formal Logic*, **30**(3), 315–342.
Warmbrōd, K. (1999). Logical constants. *Mind*, **108**(431), 503–538.
Westerståhl, D. (1989). Quantifiers in formal and natural languages. In Gabbay, D., and Guenthner, F., editors. *Handbook of philosophical logic*, Vol. IV. Dordrecht: D. Reidel, pp. 1–131.
Woods, J. (2014). Logical indefinites. *Logique et Analyse*, **227**, 277–307.
Zimmermann, T. E. (1999). Meaning postulates and the model-theoretic approach to natural language semantics. *Linguistics and Philosophy*, **22**(5), 529–561.
Zimmermann, T. E. (2011). Model-theoretic semantics. In Maienborn, C., von Heusinger, K., and Portner, P., editors. *Semantics: An International Handbook of Natural Language Meaning*, Vol. 33. Berlin: Walter de Gruyter.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF HAIFA
HAIFA, ISRAEL
E-mail: gsagi@univ.haifa.ac.il

https://doi.org/10.1017/S1755020317000247 Published online by Cambridge University Press