What Was Tarski's Thesis about Logical Truth? And Is It True?

MARIO GÓMEZ-TORRENTE
ICREA & Univ. de Barcelona, Dep. de Lógica, Baldiri Reixac, s/n, 08028 Barcelona, Spain
Instituto de Investigaciones Filosóficas, UNAM, México DF 04510, Mexico
mariogt@servidor.unam.mx

Rough draft, June 2004

The name 'Tarski's Thesis' should name a thesis by Tarski postulating the extensional equivalence between a pretheoretical concept and a more technical concept. Presumably Tarski held a thesis of this kind postulating the extensional equivalence between pretheoretical truth for a language L and the technical concept of truth for L that he showed how to define (for L a classical quantificational language). So that thesis about truth, whatever its exact form, might be called 'Tarski's Thesis'. However, my concern here is not with Tarski's theory of truth, but with his theory of logical truth and logical consequence. So I'll reserve the name 'Tarski's Thesis' for the strongest thesis of this kind that Tarski postulated about logical truth.

What was Tarski's Thesis? I'm not sure. Part of what I intend to do is to make explicit a few theses about logical truth that sound Tarskian somehow, and to offer a quick evaluation of each. Towards the end, I will do just this with a thesis that in my view most deserves the name 'Tarski's Thesis'. And for each of these theses I will claim that either it is presumably true but too weak, or it is too strong and false, or it is appropriately strong but also false. So to some extent this talk belongs to the genre of critiques of Tarski's theory of logical truth.

Here is a first candidate:

(T1) A sentence of a classical propositional/quantificational language is logically true in the pretheoretical sense iff it is true in all classical propositional/quantificational models which (re)interpret its constants (other than its classical propositional/quantificational logical constants).

(T1) is very specific about the class of sentences it talks about. It is also very specific about the class of models it talks about, and about the notion of truth in a model that is at stake, which are just the classical, Tarskian ones: in particular, a model is seen as a sequence composed of a set-domain of quantification built out of existing objects, plus extensions drawn from this domain for the predicate and individual constant letters of a language in the relevant class. And (T1) is very specific about the set of logical constants it talks about, namely: the truth-functional propositional connectives, the classical quantifiers of finite order and the predicate of identity.

Is (T1) true, or is it false? Most people until relatively recently have thought that it must be true (or at least most of those who agree that the higher-order quantifiers are logical constants have thought that it must be true). But recently several people have given arguments purporting to show that certain classical quantificational sentences are true in all classical quantificational models which (re)interpret their constants (other than their classical logical constants) and yet are not logically true. John Etchemendy and Vann McGee are perhaps the foremost examples of proponents of alleged counterexamples to (T1). The issue is a subtle one. I have talked at length about it elsewhere. My view is that at least Etchemendy's and McGee's alleged counterexamples are unconvincing, and furthermore that there are general theoretical reasons to believe that (T1) is true (reasons which are absolutely conclusive, or nearly so, when (T1) is restricted to sentences of propositional or first-order classical quantificational languages).

I cannot go into the alleged counterexamples and the general arguments supporting (T1) here without digressing excessively. But what I want to claim is not that (T1) is true (nor, of course, that it is false). I want to claim that (T1) is, although probably true, too weak to be called 'Tarski's Thesis'. I believe it's a thesis Tarski would have been ready to postulate, and in fact in the paper I just alluded to I (inadequately) refer to (T1) as 'Tarski's Thesis'. But it is not the strongest thesis Tarski would have been ready to postulate. The reason why it is too weak is that it talks about a very restricted set of logical constants. It's not that restricting one's claims in a way like this is bad in itself. What bothers me is that (T1) is certainly not the thesis that Tarski had in mind when he wrote his now classic 1936 paper on logical consequence. For Tarski was clearly not concerned with the statement of a thesis about a severely restricted set of logical constants (although he contemplated the possibility that the notion might be so hopelessly obscure as to make arbitrary any delimitation of the borderline between logical and non-logical constants). Presumably Tarski would have accepted that other constants besides the classical logical constants of quantificational languages are logical constants, even assuming that the borderline between logical and non-logical constants is not arbitrary.

This makes me think that we need something which, unlike (T1), is reasonably liberal about the class of logical constants it talks about. Let's consider this:

> (T2) A sentence of a formal language which possibly extends a classical propositional/quantificational language with new logical constants which are propositional connectives, quantifiers or predicates is logically true in the pretheoretical sense iff it is true in all classical propositional/quantificational models which (re)interpret its constants (other than its logical constants).

(T2) is just like (T1), but it does not restrict itself to any specific set of logical constants; consequently, it also does not restrict itself to sentences of classical quantificational languages, but talks about sentences with possibly new logical constants which are propositional connectives, quantifiers, and predicates having the same syntax as their analogues in classical quantificational languages. For example, one of the languages (T2) talks about is a typical quantificational modal language with identity.

One problem with (T2) is that, no matter how one understands the notion of truth in a model that appears in its formulation, and given a natural choice of logical constants, it is obviously false, and it is pretty absurd to think that Tarski might have had something like this in mind. Classical propositional or quantificational models are clearly not appropriate for a theory of the logical properties of the modal logical constants, such as $\Box$ ("necessarily"). A classical propositional or quantificational model provides only information about the extension (in the model) of the non-logical vocabulary. And we know that the satisfaction conditions of formulas of the form $\Box$A (in a world) cannot be appropriately given without reference to (at least) the extension of the non-logical vocabulary in several models (doing the job of possible worlds).

To make the problem vivid, concentrate on a simple propositional example. A classical propositional model simply assigns a truth-value to each propositional letter. Suppose M is such a model. How are we to understand the notion of truth in M for a sentence like $\Box p$ (so that $\Box$ is not subject to reinterpretation)? There are just four possibilities: (1) $\Box p$ is true in M when p is true in M and also true when p is false in M; (2) $\Box p$ is true in M when p is true in M and false in M when p is false in M; (3) $\Box p$ is false in M when p is true in M and true in M when p is false in M; (4) $\Box p$ is false in M when p is true in M and also false in M when p is false in M. Given (1) or (2), the sentence ($p \supset \Box p$) is true in all classical models, and it is thus a counterexample to (T2), for it is not an intuitive logical truth; given (3) or (4), the sentence ($p\supset\sim\Box p$) is true in all classical models, and is again a counterexample to (T2).

So we need something weaker than (T2). One possibility would be to refine (T2) by specifying a finite set of extensional and non-extensional constants we want to make our claim about and trying to be specific about some corresponding non-classical notions of model plus accompanying notions of truth in a model. For example, consider this:

> (T3) A sentence of a classical propositional/quantificational/modal language is logically true in the pretheoretical sense iff it is true in all propositional/quantificational/Kripke models which (re)interpret its constants (other than its classical propositional/quantificational/modal logical constants).

(T3) (or something a bit more precise) may well be true. In this respect it may be very much like (T1). But, as in the case of (T1), one problem with (T3) and similar theses is that they talk about a very restricted set of logical constants. In fact, as long as one restricts oneself to a smallish finite set of logical constants (as in (T1), (T3) and similar theses), even the set of *extensional* logical constants among them may, for all we know, be always too restricted. And a problem peculiar to (T3), that in any case disqualifies it as a suitable Tarskian thesis, is that it talks about a notion of model (Kripke models) that Tarski simply does not talk about or even adumbrate in his classic paper of 1936.

Another possibility is this:

> (T4) For every formal language *L* which possibly extends a classical propositional/quantificational language with new propositional connectives, quantifiers and predicates, there is a peculiar class *C* of models for the language and a notion of truth in a model for *L* and *C* such that: a sentence *S* of *L* is logically true in the pretheoretical sense iff *S* is true in all models in *C*.

(T4) says, roughly, that no matter what language of those (T2) talked about we consider, we can find some sort of not necessarily classical notion of model such that the sentences logically true in that language are precisely the sentences true in all the models of that sort. As I see it, (T4) has a problem, not with indeterminacy or with falsity but with strength. (T4) seems just too weak, because it's *trivially true*. For let *S* be a sentence of a language of the relevant kind. If *S* is not logically true in the intuitive sense, simply pick some mathematical entity $E(S)$ associated with S and say that S is false in $E(S)$; let C be the class of all these entities $E(S)$ for S not logically true; and finally say that for every S that is logically true in the intuitive sense, S is true in all the members of C.

It might be claimed that triviality is not a problem, that only falsity is. Maybe so. But this claim would only imply that (T4) is something Tarski would have been ready to accept (and how not, if it's trivially true?). It does not imply that it's the strongest claim Tarski would have been ready to postulate. Tarski quite obviously saw his thesis about logical truth as something open to refutation. And (T4) does not claim anything about the notion of a logical constant which figures so prominently in Tarski's considerations and doubts. So in any case (T4) is not Tarski's Thesis, in the sense in which I'm using this name.

To summarize, I think that the thesis that Tarski probably had in mind in 1936 was something weaker than (T2) but not as specific as (T1) or (T3) or as trivial as (T4). Further, Tarski's Thesis must have been one that seems reasonable to postulate when one restricts one's attention to classical propositional/quantificational models, for these are the models that Tarski clearly has in mind. Finally, the Thesis must make a broad claim about the notion of a logical constant.

The following thesis, (T5), satisfies all these desiderata, and seems to me quite likely to capture the essence of what Tarski had, at least implicitly, in mind:

> (T5) A sentence of a formal language which possibly extends a classical propositional/quantificational language with new *extensional* logical constants which are propositional connectives, quantifiers and predicates is logically true in the pretheoretical sense iff it is true in all classical propositional/quantificational models which (re)interpret its constants (other than its extensional logical constants).

(T5) is just like (T2), but it restricts itself to quantificational languages with extensional logical constants, and tacitly presupposes a natural extension of the classical notion of truth in a model for such languages (see below). This was very likely Tarski's intent. There are well-known dismissive remarks of Tarski about the presumable impossibility of giving non-extensional constants "any precise meaning" (these remarks are in his classic paper on truth). Besides, it's clear that if he had had in mind a thesis about non-extensional logical constants, he would not have restricted himself to extensional models such as the classical quantificational models, in view of elementary considerations such as those that show that (T2) is false.

What does 'extensional' mean, exactly? Again it's a somewhat vague notion, but the rough idea, which will be enough for my purposes, seems to be this (the extension to the polyadic cases is obvious):

a) A (monadic) connective © is extensional if whether a formula ©φ is satisfied by a valuation v of the variables at a world w is a function of whether φ is satisfied or not by v at w (a function of the "satisfaction value" of φ by v at w).
b) A (monadic) quantifier Q is extensional if whether a formula $Qx\varphi$ is satisfied by a valuation v of the variables at a world w is a function of the set of valuations which differ from v at most at 'x' and which satisfy φ at w.
c) A (monadic) predicate P is extensional if the satisfaction of a formula Pt by a valuation v of the variables at a world w is a function of the extension of t under v at w.

Is (T5) true? There are a number of examples in the literature that, if they were convincing, would show that (T5) is false. Two of these examples are due again to John Etchemendy and Vann McGee. Etchemendy has noted that, if one takes as logical constants the extensional monadic predicates 'P' and 'M', meaning respectively "is or was a president of the U.S. in or before 2003" and "is a male", then the quantificational sentence

(1) $(\forall x)(P(x) \supset M(x))$

is true in all classical quantificational models, since no matter what model we choose (1) will be true in the model. In this case, that means that no matter what (set-sized) quantifier domain of existent things we choose, every object in that domain will be either a non-president or a male (there haven't been any female presidents in the actual world). However, (1) is not intuitively logically true, for it is not even necessary. Or, at least, (1) is not necessary if the proposition it expresses quantifies over any one of a wide class of natural ranges for its quantifier; for example, if it ranges over "absolutely everything", or over the set of humans, etc., then (1) is not intuitively necessary, and hence it is not intuitively logically true.

Vann McGee has given another purported counterexample (but he has not categorically asserted that it is a counterexample). Take as a logical constant the extensional quantifier '$(\exists^{PC}x)$', meaning "there are at least a proper class of x's such that"; then the quantificational sentence

(2) $\sim(\exists^{PC}x)(x=x)$

is true in all classical quantificational models, since no matter what set-sized quantifier domain we choose, the sentence will be true in that domain. And yet (2) is not true for proper class-sized domains, much less analytic or necessary, and so intuitively not logically true. Or, at least, (2) is not analytic or necessary when the proposition it expresses quantifies over, e.g., "absolutely everything", or over the class of sets, etc.

But a problem for anyone who wants to use these examples against (T5) is that the arguments needed for this must be premised on suspicious choices of expressions as logical constants. There is no intuition, I think, that either 'P' or 'M' or '$(\exists^{PC}x)$' are logical constants, so the persuasive force of Etchemendy's and McGee's examples is limited.

As I implied at the beginning, I think that (T5) is false. But I hope the reason why I think it's false is more persuasive. I will argue that (T5) is false by exhibiting a certain quantificational sentence containing only constants that are intuitively logical and extensional, which is true in all classical quantificational models, but which, like Etchemendy's and McGee's sentences, is *not necessary* (in the sense that it can be used, and indeed it's at least typically used, to express a contingent proposition). Assuming that contingent sentences are not logically true, it will follow that (T5) is false. (To be sure, Tarski might not have been very impressed by a refutation of (T5) that appeals to intuitions about modality, since he was very dismissive of modal notions. However, I want to evaluate (T5) not by Tarski's criteria, but by generally accepted criteria which happen to coincide approximately with my own.)

Suppose that we extend a typical first-order quantificational language with only one new logical constant, a monadic predicate 'E' meaning "exists". A primitive predicate with this intended meaning is taken as a logical constant in treatments of quantified modal logic and of intensional logic generally (see, e.g., Kaplan's logic of demonstratives). Its intended meaning, a bit more explicitly, is given by the principle that it is to be satisfied by an object at a world if that object exists at that world. This is the common sense of existence according to which I exist in the current circumstance but I would not have existed if my parents had not met. It can be given a simple satisfaction clause in the definition of satisfaction in a classical quantificational model, very much like the clause for identity: a model *M* plus valuation *v* satisfies E(x) if *v*(x) exists. Clearly 'E' is an extensional predicate in the intuitive sense above. (In the sense that when t and t' are terms with a shared reference under a valuation *v* at a world *w*, E(t) is satisfied by *v* at *w* iff E(t') is satisfied by *v* at *w*.)

Is the predicate 'E' really a logical constant? As I just said, it has been taken to be such in intensional logic. But furthermore 'E' is certainly topic-neutral. Moreover, it (or its correlate in natural language) seems to be widely applied. Further, 'E' is a logical constant in the technical senses defined by Tarski, Vann McGee and Solomon Feferman. It is surely invariant under permutations of a model (Tarski) and even under bijections of models (with a domain of existent things), for its extension in any model (with a domain of existent things) is the full domain of the model; and it follows from its meaning that it is invariant under bijections of models (with a domain of existent things) (McGee); it is also invariant under homomorphisms of models (with a domain of existent things) in the sense recently defined by Feferman. It is not a logical constant in the technical senses of Timothy McCarthy and Gila Sher, since it is not invariant under bijections of models with a domain of possibly non-existent things; but this would seem one more defect of McCarthy's and Sher's proposals rather than a virtue.

Besides the natural assumption that 'E' is a logical constant, the only other remark that needs to be made explicit before I exhibit the sentence I promised is a remark about the evaluation at a world *w* of the proposition expressed by a formula dominated by a classical universal quantifier (a formula of the form $(\forall x)\varphi$). There are basically two kinds of propositions that such formulas might be taken to express, which are reflected in the two most standard semantics for quantificational modal logic. The first kind of proposition is a proposition the content of whose quantifier gets specified when one specifies a quantificational domain, given purely in extension, plus a property which further restricts the range of the quantifier at a particular world. Thus, for example, given this view, in order to specify the content of the quantifier in the proposition expressed by (a use of) the sentence (1), what one has to do is to specify a (nonempty) class of objects, given purely in extension, plus a property. For example, one specifies the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} plus the property of being Texan. Assuming that the other expressions have their intuitive meaning, (1) then expresses the proposition that, roughly, "each of the Texans in the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} is, if a president, a male". This proposition is true in those worlds where, like in the present one, the Texan presidents in the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} are all males. The way this idea is reflected in one of the standard semantics for quantified modal logic is as follows: given a previously specified domain D and a property Π, one says that $(\forall x)\varphi$ is satisfied by a valuation *v* at a world *w* iff every valuation *u* of the variables (with objects of the previously given domain D) which differs from *v* at most at 'x' and which assigns to 'x' an element of D that has the property Π in *w* satisfies φ at *w*. (Typically Π is taken to be the property of existence.) This clause has the effect that with a quantifier $(\forall x)$ one quantifies in a world *w* only over the objects from D that have the property Π in *w*.

The second kind of proposition that a quantificational formula might be taken to express is a proposition the content of whose quantifier gets specified when one specifies simply a quantificational domain, given purely in extension, which constitutes the range of the quantifier at any particular world. Given this view, in order to specify the content of the quantifier in the proposition expressed by (a use of) the sentence (1), what one has to do is simply to specify a (nonempty) class of objects, given purely in extension. For example, one specifies the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} as before, and that's enough. Assuming the other expressions have their intuitive meaning, (1) then expresses the proposition that, roughly, "each of the things in the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} is, if a president, a male". This proposition is true in those worlds where, like in the present one, the presidents in the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush} are all males; and it will be false, for example, in worlds where Eleanor (is still a woman but) becomes president. The way this idea is reflected in the other standard semantics for quantified modal logic is as follows: $(\forall x)\varphi$ is said to be satisfied at a world *w* by a valuation *v* of the variables with objects of the previously given domain D iff every valuation *u* of the variables with objects of D which differs from v at most at 'x' satisfies φ at w. This clause has the effect that with a quantifier $(\forall x)$ one quantifies in a world w over all the objects of the previously specified domain D.

It is natural to take the universal quantifier in both of these senses as a logical constant. Or, in the case of the first sense, it is natural to take it as a logical constant when the property Π is a logical property like existence. And in fact the two kinds of quantifiers are taken as logical constants in quantificational modal logic. Further, in both senses the universal quantifier is an extensional quantifier in the sense above (or rather, in the case of the first sense, it is extensional provided the property Π is extensional). I will from now on take the universal quantifier in the second sense just described, that in which with it one quantifies in every world over all the objects of the domain that serves to interpret the quantifier. Since there will be no property Π to worry about, the universal quantifier taken in this sense is a logical constant without qualification.

Note that under this way of understanding the content of the universal quantifier, when we consider whether, e.g., the quantificational sentence (1) is true at a world, we always ask ourselves whether each of the objects of a previously fixed domain D is either out of the extension of 'P' for that world or in the extension of 'M' for that world. Assuming that these extensions are the sets of presidents and males, respectively, what we always ask ourselves is whether each of *those same objects, the objects of that same domain* D, is either a non-president or a male in the world at issue.

Now consider the following quantificational formula:

(3) $(\forall x)E(x)$

This formula is a model-theoretic logical truth if 'E' is a logical constant, since no matter what classical quantificational model we choose (3) will be true in the model. I.e., in this case, what that means is that no matter what actual (set-sized) quantifier domain (of existent things) we choose, every object in that domain will be an existent thing (in our world). This is so at least under the usual understanding of quantificational model theory, according to which no unactualized *possibilia* form part of the domains of models. (Etchemendy's example, as well as most discussions of these issues, are based on this assumption; notice that there may well be a merely possible person who is president of the

United States and a female, even in our world—perhaps some fictional characters are *possibilia* with this property, but I do not commit myself to this.)

Is (3) logically true from an intuitive point of view? It seems not. Suppose that we specify the proposition expressed by our use of (3) by fixing the domain of its variables to be again the set {Franklin Roosevelt, Eleanor Roosevelt, George Bush, Laura Bush}. It appears that the proposition then expressed by (3) is contingent: for example, in a possible world in which Franklin's parents had not met, the proposition expressed by (3) would not be true, because at least one of the people (3) quantifies over would not exist in that world. (For related reasons, Kaplan argued that 'I exist' is a (model-theoretic) logical truth which is not necessary. For our purposes this is irrelevant, since 'I', unlike the quantifiers, is not a logical constant.)

So under all our assumptions (3) is true in all classical quantificational models but the proposition it expresses (in one use, and in many others) is not necessary. Given the uncontroversial principle about logically true formulas that any proposition they may come to express as a result of specifying the range of their quantifiers ought to be necessary, it follows that thesis (T5), what I think was Tarski's Thesis, is false.