TARSKI'S 1936 ACCOUNT OF LOGICAL CONSEQUENCE

CRAIG N. BACH

Department of Philosophy
Michigan State University
East Lansing, MI 48824-1032, USA
email: bach@pilot.msu.edu

Much has been made in the recent literature of Alfred Tarski's seminal 12-page paper, "On the Concept of Logical Consequence" [Tarski *1936*]. However, there are two puzzling aspects of Tarski's paper which thus far have avoided adequate explanation: (1) Tarski's claim that his semantic account of consequence captures *logical* inferences (involving ω-incomplete theories and Gödel sentences) that the syntactic account does not; and (2) Tarski's seemingly false claim that in a language consisting of purely logical terms the relation of formal consequence coincides with that of material consequence. The resolution of these conundrums requires a clear understanding of the differences between the current model-theoretic project and that developed by Tarski in the 1930s, in particular, with respect to the notion of logical form.

That this notion is of central importance in determining the scope and definition of logic is clearly stated by Russell in 1903. His remark helps point the way towards an exegesis of Tarski's early work and motivates the present discussion:

> I confess, however, that I am unable to give any clear account of what is meant by saying that a proposition is "true in virtue of its form." But this phrase, inadequate as it is, points, I think, to the problem which must be solved if an adequate definition of logic is to be found. [Russell *1903*, xii]

The notion of logical form plays a key role in Tarski's expressed goal to define the proper concept of logical (read formal) consequence. I will argue that in his 1936 paper, Tarski employed an expanded notion of formality, distinct from the one in current usage, in developing and supporting his original model-theoretic definitions. Once I have outlined Tarski's logical program, I will be able to show, regarding (1), why Tarski considered his semantic account to be successful in cases where the syntactic account fails and, regarding 2), that within the logical framework set out by Tarski, the relation of formal consequence does reduce to that of material consequence for languages with no nonlogical vocabulary.

1. The Inadequacy of the Syntactic Account. Prior to Tarski's 1936 paper, the concept of logical consequence had been defined using purely syntactic methods. As Tarski [*1936*; *1956*, 410] puts it, "logicians thought that these few rules of inference exhausted the content of the concept of consequence." In his 1931 paper on truth and in his 1933 paper on ω-consistency, Tarski himself defines the consequence relation in syntactic terms. He defines the set of logical consequences of a given axiom system as the closure of the axiom system under specified rules of inference. However, by 1933 Tarski noticed a deficiency in the syntactic account, and in his 1936 paper proceeded to give an alternative, semantic account of logical consequence. This move was preceded by two major results in the field; that of the existence of ω-incomplete theories [Gödel *1931*], [Tarski *1933*] and Gödel's general incompleteness results [Gödel *1931*].

The ω-incompleteness result laid out in Tarski's 1933 paper is given with reference to a particular theory, say T, with the following properties:

$T \vdash A_0$: where A_0 states that 0 possesses a given property A.

$T \vdash A_1$: where A_1 states that 1 possesses a given property A.

.

.

.

$T \vdash A_i$: for every natural number i.

However, it is not the case that:

$T \vdash (\forall n)\, A_n$: where $(\forall n)\, A_n$ states that every natural number possesses the property A.

The syntactic account fails in this case to capture the intuitive result indicated. Tarski makes this point at the end of his 1933 article on ω-incompleteness:

> Formerly it could be assumed that the formalized concept of consequence coincides in extension with that concept in everyday language, or **at least that all purely structural operations, which unconditionally lead from true statements to true statements**, could be reduced without exception to the rules of inference employed in the deductive disciplines. It might also be

> thought that the consistency of a deductive system is in itself a sufficient guarantee against the appearance of statements in the system which — **on account of their mutual, structural relations** — cannot both be true. Since, however, there are systems which are on the one hand ω-incomplete and on the other consistent, but not ω-consistent, the basis of both these assumptions is removed. [Tarski *1933*; q.v. *1956*, 294][1]

Again Tarski makes the same point in his 1936 article on logical consequence, explicitly noting that the syntactic account is not capable of defining the notion of logical consequence in its entirety:

> It [an ω-incomplete theory] shows that the formalized concept of consequence, as it is generally used by mathematical logicians, by no means coincides with the common concept. Yet intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots$. Provided all these statements are true, the sentence A must also be true. [Tarski *1936*; q.v. *1956*, 411]

The inability of the syntactic account of consequence to capture intuitively valid argument forms, however, is not limited to a particular theory, axiomatization, or selection of rules of inference. Gödel's results show that for any consistent theory of a minimal complexity with a recursive set of axioms there will be a sentence which follows in an intuitive sense, but which is not provable from the given deductive theory. Tarski noted this as follows:

> In every deductive theory (apart from certain theories of a particularly elementary nature), however much we supplement the rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of inference. [Tarski *1936*; *1956*, 412]

Tarski considered both the universal sentence in the ω-incompleteness example and the Gödel sentence to (1) follow in the usual sense from, (2) follow by purely structural operations from and (3) to be logical consequences of the theories in which they were formulated.

Two conditions are thereby set on the development of an accurate definition of logical consequence. The definition must adhere to the

---

[1]It is important to note that Tarski here considered that the statement $(\forall n)A_n$ follows from the collection of A_n's by "purely structural operations" (i.e., follows formally). Boldface mine.

common usage of the concept and it must capture those inferences which are based on purely structural operations. The syntactic account fails to provide such a definition.

2. Tarski's 1936 Definition of a Model. According to Tarski, the definition of logical consequence must be given in terms other than those which are purely syntactical: "In order to obtain the proper concept of consequence, which is close in essentials to the common concept, we must resort to quite different methods and apply a quite different conceptual apparatus in defining it" [Tarski *1936*; *1956*, 413]. To this end, Tarski applies the tools and methods he had developed in his earlier paper on truth to the definition of logical consequence.[2] The most important notion taken from that paper and used in his definition is that of *satisfaction*, which gives rise to the related notion of *model*. Tarski's definition of logical consequence runs as follows:

> **Definition T-1**: The sentence X follows logically from the sentences of the class K if and only if every model of the class K is also a model of the sentence X.

On the face of it, it seems that this definition is exactly the one used in most modern textbooks in logic. However, there are critical differences.

One difference between the current definition and the one Tarski put forth in 1936 involves the definition of a logical model. Tarski defines a model as follows:[3]

Given:

[1] A language which contains variables corresponding to each extra-logical constant in the language.

---

[2] I am taking the period, roughly between 1929 and 1936, as comprising a distinct intellectual period in the work of Alfred Tarski. After this time certain key elements of his work change and become what is today the received view. In "The Concept of Truth and Formalized Languages," Tarski points this out:

> In the course of the years 1929 to 1935, in which I reached the final definition of the concept of truth and most of the remaining results described here, and in the last year of which the whole work appeared for the first time in a universal language, the questions here discussed have been treated several times. [Tarski *1931*; q.v. *1956*, 277]

Thus, there are important conceptual links between the 1936 paper on logical consequence and the 1931 paper on truth.

[3] I will here forego defining the standard notion of a model as I am assuming most readers are familiar with it.

[2] A set of sentences, say K, of that language.

[3] A set of sentential functions K' obtained from K by replacing within each sentence in K all occurrences of extra-logical constants contained in the sentence by their corresponding variables.

Then:

An arbitrary sequence of objects which satisfies every sentential function in K' is said to be a *model* of K.

To differentiate between this formulation and the standard one, the following Tarskian approach to an example from the first-order theory of the calculus of classes will help (this example is a bit awkward and quite limited, but will serve our present purpose well):

Let

$K = \{a_1 \subseteq a_2, a_2 \subseteq a_3\}$,

$X = a_1 \subseteq a_3$, where a_1, a_2, a_3 are names of classes of individuals.

If all occurrences of non-logical constants are replaced by their corresponding variables, we arrive at the following set of sentential functions K' and sentential function X':

$$K' = \{x_1 \subseteq x_2,\ x_2 \subseteq x_3\},$$

$$X' = x_1 \subseteq x_3.$$

Any sequence of classes which satisfies the set of sentential functions K' will satisfy the sentential function X'. Therefore, the sentence is determined by the analysis to be a logical consequence of the set of sentences K.

From the perspective of the standard model-theoretic account, this definition equates logical consequence with the truth of the universal generalization of the conditional whose antecedent is the conjunction of the sentences in K, and whose consequent is the sentence X, in a standard first-order model. In other words, saying that every sequence which satisfies the sentences in K' satisfies the sentence X' is to say that the first-order sentence:

$$(\forall x_1)(\forall x_2)(\forall x_3)\ ((x_1 \subseteq x_2 \wedge\ x_2 \subseteq x_3) \rightarrow x_1 \subseteq x_3)$$

is true in the intended interpretation.

In this example, the meanings of the standard logical constants of the language are held fixed, as well as the meaning of the '$\subseteq$' symbol. Also, the sequences that do or do not satisfy the formulas range over, and differ by, only the classes which comprise them. The domain of discourse is missing entirely from the discussion of models in the 1936 paper; it is not given a place in the formulation of sequences which Tarski defines as models. This leads to the relation of logical consequence boiling down to the truth of a universal sentence in a fixed domain.[4]

3. The Domain of Discourse. As was noted above, the domain of discourse is not explicitly stated in Tarski's 1936 formulation of a logical model. There is no parameter in the sequences, that Tarski defines as models, with which to relativize the elements of the sequence to a particular domain. However, Tarski's 1936 definition of a logical model can easily be amended to bring the definition in line with contemporary usage: one need only add a parameter to the sequences which relativizes the domain from which the elements of the sequence are taken. Why is it, then, that no mention of the domain of discourse is made in the 1936 article?

The problem of the role of the domain in Tarski's 1936 analysis has been pointed out several times in the recent literature. In his review of Etchemendy's book, Vann McGee remarks that, "In particular, Tarski's original analysis [of logical truth] makes no provision for the special role of the universe of discourse, so that it gives a faulty account of the quantifiers" [McGee *1992*, 273]. Hodges [*1986*] has claimed that Tarski merely left out discussion of domains due to the technical abilities of the audience to which he presented the 1936 paper — a point also made in Carnap's autobiography [Carnap *1963*, 61–62]. However, there are reasons to think that Tarski, when formulating his definition of logical consequence, was not merely neglectful of the role of the domain. He was well aware of the role of the domain in defining truth, and the different results that can be obtained when the domain of discourse is altered. However, he set those matters aside from his main task. To see this we need to turn to Tarski's 1931 paper on truth.

[4]It is quite straightforward to use the above formulation to define a model in the manner of contemporary usage: one need only fix a particular domain to each sequence. The objects which comprise the sequence would then only be drawn from the given domain. Thus a first-order model is defined as an infinite sequence S with a built-in parameter for the domain of discourse:

$$S = \langle a_1, a_2, \ldots, a_n, \ldots \rangle \text{ s.t. } a_i \in |U|, \text{ where } |U| \text{ is a domain-set.}$$

This is, in its essentials (modulo a discussion of what terms are held fixed in the language), the contemporary definition of a logical model.

In that paper, Tarski takes up the calculus of classes as an example with which to set out his semantic definition of truth.[5]

In the monograph on truth, Tarski makes the distinction between truth *simpliciter* and the domain-relative variety explicit. In the third section of that paper, he defines truth for a sentence of the language of the calculus of classes. As mentioned above, the domain of discourse in this discussion is held fixed — the domain is taken to consist of an infinite number of individuals. The truth of a particular sentence from the formal language is determined within this infinite domain. However, he is also well aware of results relative to domain size:

> In the investigations which are in progress at the present day in the methodology of the deductive sciences . . . another concept of a relative character plays a much greater part than the absolute concept of truth and includes it as a special case. This is the concept of correct sentence in an individual domain *A*. By this is meant (quite generally speaking) every sentence which is true in the usual sense if we restrict the extension of individuals considered to a given class *A*, or — somewhat more precisely — when we agree to interpret the terms 'individual', 'class of individuals', etc. . . . , as 'element of the class *A*', 'subclass of the class *A*', etc . . . respectively. [Tarski *1931*; q.v. *1956*, 199]

Again, the point is echoed in the following quote:

> As derived concepts we introduce the notion of a correct sentence in an individual domain with *k*-elements and the notion of a correct sentence in every individual domain (generally valid). [Tarski *1931*; 1956, 200]

He even proves a series of domain-relative theorems, including a version of the Löwenheim-Skolem theorem.

It is clear that Tarski is aware of the role that the domain played in these results — it is not, however, his main goal in the paper to take up

---

[5]There is a small technical peculiarity in the account which does not directly affect our present concerns, but which is nevertheless worth mentioning. The domain in the example is not only held fixed, but is once removed from the role it plays in the standard account. The objects under discussion are classes, and all sequences of objects involved in the definition of satisfaction are sequences of classes. This means that all first-order quantifiers range over classes of individuals. The standard account would take the domain of discourse to be a set of classes —in Tarski's view, the set of all classes considered in the standard interpretation and given by the assumptions put forth in the meta-theory. However, Tarski pushes the domain further back by taking it to be a set of individuals out of which the classes are formed.

those considerations. Tarski's goal in the first section of the paper is to define *truth* for the calculus of classes: "The problem still remains of clarifying the relation of the absolute concept of truth defined in Definition 23 to the concepts we have just investigated" [Tarski *1931*, 207]. Again, hopefully without belaboring the point, this distinction is brought out in the statement of Theorem 26 from Tarski's truth paper:

> **Theorem 26**: If A is the class of *all*[6] individuals then $x \in Tr$ iff x is a correct sentence in the domain A; thus if κ is the cardinal number of the class A, then $Tr = Ct_{\kappa}$.

Tarski makes a clear and pronounced distinction between the notion of truth *simpliciter*, and the notion of truth in a domain of such-and-such a size. As indicated by the number of remarks made in the truth paper, it is also clear that Tarski considers the clarification of this distinction to be quite important.

In setting out his definition of truth (the absolute variety), Tarski holds fixed the domain of individuals corresponding to the language under discussion. This contention also is supported by Tarski's claim that his semantic definition of logical consequence captures the universal sentence discussed in the example of an ω-incomplete theory. The existence of ω-incomplete theories requires the domain-set to contain both non-standard elements and the natural numbers as a proper subset (up to isomorphism). Therefore, to capture the types of arithmetic inferences Tarski claims that his semantic account does capture, the domain-set must be restricted to the natural numbers.[7] This entails that assumptions made about the fragment of mathematics under discussion (e.g., the calculus of classes, arithmetic) will determine the size of the domain[8] and the meaning of various symbols in the language.

---

[6]Italics mine. Ct_k refers to the set of correct sentences (true) for a domain of size k.

[7]This point is made quite succinctly by Henkin: "Tarski, and later Gödel, showed the existence of consistent systems which were ω-inconsistent. We can now see that such systems can and must be interpreted as referring to a non-standard number system whose elements include the natural numbers as a proper subset" [Henkin *1950*, 91]. At first blush, it may seem circular to use the very claim I am analyzing as support for the analysis. However, to show that the claim entails certain properties of the doman-set for which I have provided independent textual and conceptual evidence is not circular.

[8]This claim is supported as well by the following quote from the truth paper: "Because we can show on the basis of the system of assumptions here adopted, that the class of all individuals is infinite, Th. 26 in combination with Th. 12

A remark, then a question. Tarski clearly knew of the work of the Göttingen school concerning domain-relative results, and he was well aware of the role that the domain of discourse played in the set-theoretical, semantic definition of truth. Why then did he not use this knowledge when he formulated his definition of logical consequence? I will attempt an answer to this question after covering a little more ground in the following sections.

4. "Following in the Usual Sense": Ordinary Consequence and Natural Languages. In his 1936 paper, Tarski makes several references to an *ordinary* consequence relation. These references are important to the present discussion in that their misinterpretation has lead some commentators (most notably Etchemendy) to incorrectly assess both the types of languages that Tarski considered amenable to his definition of consequence and the aspects of those languages central to the Tarskian analysis. Inaccurate determinations of these types directly lend themselves to an inaccurate evaluation of the Tarskian analysis in general. If it is claimed that the Tarskian account was intended to capture the ordinary concept of consequence, and by this it is meant that the account was intended to capture all necessary inferences irrespective of the language in which these inferences are made, then the Tarskian analysis certainly fails. Therefore, it will do the discussion well to take a careful look at Tarski's references to the *ordinary* consequence relation.

Tarski, on several occasions, states that the task of providing an analysis of logical consequence should take into account the ordinary usage of that term:

> The concept of logical consequence is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision, . . . ; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life. [Tarski *1936*; *1956*, 409]

There is ample textual evidence to support the claim that the intent of the Tarskian account of logical consequence was to capture a formal notion which was "close in essentials" to the common usage of that term.

Etchemendy takes these passages from Tarski's 1936 paper to support the claim that the most important concept underpinning the notion of logical consequence which Tarski was attempting to capture is the modal relationship of necessity. His point is that if one is attempting to capture the *ordinary* notion of consequence — the one used in the work-a-day world

---

makes a structural characterization of true sentences [for the language of the calculus of classes] possible" [Tarski *1931*; q.v. *1956*, 208].

— then one is attempting to capture the notion of necessary consequence, and the most important feature of that relation is the modal relationship of necessity. Simply put: If an argument is valid (the argument's conclusion is a logical consequence of its premises) and the premises of the argument are all true, then the conclusion *must* be true:

> That this is the single most prominent feature of the consequence relation, or at any rate of our ordinary understanding of that relation, is clear from even the most cursory survey of texts on the subject. [Etchemendy *1990*, 81]

At first glance, this remark seems rather innocuous, but on closer inspection of the inferences which Etchemendy draws from this remark, it can be seen that something is missing from his account of Tarski's analysis.

The above passages are used by Etchemendy to support two main aspects of his attack on the Tarskian conception of logical consequence. The first is the emphasis placed on the modal aspect of the relation of logical consequence, leaving out almost entirely the idea that logic addresses the formal nature of argumentation as well. The second use that Etchemendy makes of the above passages is to introduce a wide range of natural language examples. It seems that the ordinary concept of consequence Tarski was attempting to capture with his analysis is taken by Etchemendy to be the consequence relation as it is used in ordinary, natural languages.

A proper reading of the relevant passages from Tarski's 1936 paper and surrounding texts shows that these passages do not lend support to either an exclusive concentration on the modal aspects of the consequence relation, or to the introduction of the Etchemendian type natural language examples. As was pointed out at the beginning of the paper, there are two important conditions placed on a satisfactory definition of logical consequence: that it conform to common usage (modal aspect) and that it capture inferences based on purely structural operations (formal aspect). To obtain a better idea concerning exactly how the notion of 'ordinary consequence' plays itself out in the Tarskian analysis, I will need to turn elsewhere in the 1936 paper, and back again to Tarski's 1931 monograph on truth. In Tarski's paper on truth, explicit mention is made of the limitations of a formal analysis: "A thorough analysis of the meaning current in everyday life of the term *true* is not intended here" [Tarski *1931*, 153]. A similar sentiment is expressed later in the same paper: "The attempt to set up a structural definition of the term true sentence — applicable to colloquial language — is confronted with insuperable difficulties" [Tarski *1931*; *1956* 164)]. It is easy to see that these remarks have a direct bearing on the 1936 paper as well. The Tarskian account of logical consequence is dependent on there being a formal, semantic definition of truth for the language being studied. Thus, any discussion of the limitations of an analysis of truth carries over to the analysis of logical consequence as well.

Even in the 1936 paper itself a reference is made to the limitations of the formal analysis: "Every precise definition of this concept [logical consequence] will show arbitrary features to a greater or lesser degree" [Tarski *1936*; *1956*, 409]. The references to the ordinary concept of consequence do not show that Tarski was attempting to capture that notion as it is used in natural language argumentation, with all of its inherent vagueness and ambiguity. The strict formal guidelines placed on such an analysis in the truth paper apply to the discussion in the 1936 paper as well. Both accounts require that the languages under study are of a precise formal nature.

The exclusive position given the modal relation of necessity in Etchemendy's account of logical consequence is done so by removing discussion of the condition of formality on that relation. As briefly indicated in the introduction above, the notion of form is central to the discussion of the concept of logical consequence. It also takes up a central position in Tarski's early work. This is indicated in a passage directly from the 1936 paper:

> [W]e are here concerned with the concept of logical, i.e. formal consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, (i.e. objects of reference). [Tarski *1936*; *1956*, 414]

Sher takes up this point as well, in her recent book, making the formality condition explicit in the definition of logical consequence:

> (C2) Not all necessary consequences fall under the concept of logical consequence: only those in which the consequence relation between a set of sentences *K* and a sentence *X* is based on formal relationships between the sentences *K* and *X* do. [Sher *1991*, 40]

That Etchemendy says nothing concerning this second condition on the relation of logical consequence calls into question his reading of the relevant passages.[9]

---

[9] There is one other point involving the translation of the 1936 paper on logical consequence that may lend support to the Etchemendian reading. When discussing the example of an ω-incomplete theory, Tarski remarks that, "it seems certain that the universal sentence *A* follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$" [Tarski *1936*; q.v. *1956*, 411]. This remark, as presented in [Tarski, 1954], is translated from the German text. In the German version of the paper, the term *inhaltlich* is used in place of the term usual in the above quotation. However, *inhaltlich* carries with it certain

Tarski's stated intention of providing an account of logical consequence that is "close in essentials" to ordinary usage does not commit him to supplying an account of logical consequence which captures all the subtleties of that relation as it is used in natural languages, or to one which captures the more encompassing relation of necessary consequence. The modal aspect of the relation provides only half of the story; the Tarskian analysis of logical consequence, in its essentials, is dependent upon the notion of form and how that notion is explicated.

5. The Relation Between Formal and Material Consequence. In the 1936 paper, Tarski makes the following claim: "In the extreme case we could regard all terms of the language as logical. The concept of formal consequence would then coincide with that of material consequence" [Tarski, *1956*, 419]. In response to this remark, Sher offers the following problematic example [Sher *1991*]. Consider the following two sentences:

$$\varphi : (\exists x)(\forall y)(x = y)$$

$$\theta : (\exists x)(\exists y)(x \neq y \wedge (\forall z)(z = y \vee z = x))$$

According to the standard account, there are no non-logical symbols in the above sentences — the meaning of every symbol is held fixed. However, it may well be that θ is a material consequence of φ, but it is not the case that θ is a logical consequence of φ — consider a logical model whose domain has one element. At first blush, Tarski's remark seems to be just plain false.

To address Tarski's claim, it is necessary to first give a more detailed account of the types of languages Tarski thought were amenable to his analysis. In the introduction to his 1931 paper on truth, Tarski sets his project aside from that of the Hilbert school. Tarski's intended goal is to

---

connotations relevant to the present discussion, which the term usual does not. The passage in German is better translated as, ". . . *A* follows *with respect to content* from the . . . ." It may be argued that the German translation points to the fact that Tarski had in mind a consequence relation more concerned with the content of sentences than is the standard account, and that this fact may support the Etchemendian concentration on the modal notion of necessity in the paper. However, the point is not strongly supported on these grounds. If attention is drawn to the original Polish, the relevant term used is *intuicyj*, which is best translated into English as *intuitively*, or as "in common, everyday usage." Thus, the Polish term is more in line with the English term usual, as it is used in the above quotation.

consider formal languages which are both precise and meaningful, i.e., interpreted. This is made clear in two remarks he makes in the truth paper:

> It remains perhaps to add that we are not interested here in "formal" languages and sciences in one special sense of the word "formal," namely sciences to the signs of which no material sense is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful. We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider. [Tarski *1931*; *1956*, 166]

and again:

> The expressions which we call sentences still remain sentences after the signs which occur in them have been translated into colloquial language. The sentences which are distinguished as axioms seem to us to be materially true, and in choosing rules of inference we are always guided by the principle that when such rules are applied to the true sentences the sentences obtained by their use should also be true. [Tarski *1931*; *1956*, 167].

These remarks are similar to remarks given by Frege in describing his own work in the *Begriffschrift*: "My intention was not to represent an abstract logic in formulas, but to express a content through written signs in a more precise way than it is possible to do through words" [Frege *1883*, 1]. The stated goal is not one of complete formalization at the cost of presenting a meaningful language, but rather to give a formally, precise explication of a fully meaningful language.

To provide an account of how the meanings of the logical quantifiers are held fixed in such an interpreted language, it will be helpful to first make a distinction between the semantic role that the overall system of models plays in the analysis and the semantic contribution of each particular model.[10] In terms of modern usage, the logical quantifiers have a fixed meaning across all models; the universal quantifier is assigned the domain-set and the existential quantifier is assigned non-empty subsets of the domain-set. However, there is another sense in which the meanings of the quantifiers can be held fixed. This can be done by taking the meanings of the quantifiers to be their assignment under each particular model. On this second interpretation, each assignment of a domain-set changes the meaning of the quantifiers. Therefore, to say that the meanings of the logical quantifiers are fixed is to say that the assignment of the domain-set is fixed.

Tarski did not utilize purely formal languages in his discussion of truth and logical consequence; rather he concerned himself with interpreted formal

[10] A similar distinction is made in [Sher *1991*].

languages. In these languages all of the symbols in the language have a precise meaning independent of the model-theoretic apparatus, including what and how many objects are quantified over; the size of the domain is fixed by the underlying assumptions inherent in the theory under investigation. For example, in the calculus of classes these assumptions necessitated the domain size to be infinite and in the language of arithmetic that the doman be restricted to the natural numbers. The model-theoretic apparatus is set up to reflect these meanings. On this reading, the meanings of the quantifiers are held fixed in the second sense discussed above. The "counter-example" presented by Sher to Tarski's remark concerning the reduction of formal consequence to material consequence does not provide an example which shows the remark to be false — it is not applicable to the 1936 project.[11]

These remarks raise serious questions. Most importantly, how is the notion of formality being used by Tarski to insure that his definition of logical consequence reflects structural features of the language and insures that the modality condition on that relation is still met? More precisely, the concern is that if the relation of logical consequence can be reduced to material consequence, then his definition does not retain the necessary modal features that he was trying to capture. The answer to these questions involves the close relationship in Tarski's work between formal and mathematical inference, and awaits the discussion of Carnap given in the next section.

6. Carnap's "Logical Syntax of Language". Some insight into the exact nature of Tarski's 1936 definition of logical consequence can be gleaned from the writings of Carnap, in particular, from his *Logical Syntax of Language*. Throughout the 1936 paper, Tarski makes reference to Carnap's work. He states that Carnap's work, like his own, intends to set out the proper concept of consequence which is at odds with the received

[11]An interesting question has been raised by Gila Sher in response to this interpretation. Consider two languages L_1 and L_2 each with distinct ontologies O_1 and O_2, respectively, such that upon the removal of all non-logical constants the languages are identical. Call the languages without non-logical terms L_1' and L_2', respectively. It is clear that L_1' and L_2' are identical. The question is what ontology do we assign to the symbols of L_1' and L_2'; O_1 or O_2? In response, note that the question assumes that the languages L_1' and L_2' are purely formal, in particular with respect to the meaning assigned to the logical quantifers. However, according to the present reading, the logical quantifers are held fixed over a specific domain-set and their meaning would not change by removing all of the non-logical constants (see the example of the ω-incompleteness discussed earlier in the paper).

syntactic account. Tarski even goes as far to note that his semantic definition of consequence is provably equivalent to Carnap's definition: "On the basis of all these conventions [translational conventions between the two definitions] and assumptions it is easy to prove the *equivalence of the two definitions*" [Tarski *1936*; *1956*, 418].

Tarski's and Carnap's discussions on the concept of logical consequence are also similar in structure and in scope. In *Logical Syntax*, Carnap, like Tarski, is concerned with formal relations in his discussion of logical consequence:

> We shall see that the logical characteristics of sentences (for instance, whether a sentence is analytic, synthetic, or contradictory; . . .) and the logical relations between them . . . are solely dependent upon the syntactical structure of the sentences. [Carnap *1937*, 1–2].

Both thinkers are interested in exploring the structural relations between sentences for a very precise range of formal languages.

Carnap, as was Tarski, also is concerned with the inadequacy of the derivational account of the concept of consequence. The following remark from *Logical Syntax* makes this clear: "It is impossible by the aid of simple methods to frame a definition for the term consequence in its full comprehension. Such a definition has never yet been achieved in modern logic (nor, of course, in the older logic)" [Carnap, *1937*, 27]. Again, this point is made later in the same work:

> The term 'derivable' is a narrower one than the term 'consequence'. The latter is the only one that exactly corresponds to what we mean when we say: "This sentence follows (logically) from that one," or: "If this sentence is true, then (on logical grounds) that one is also true". [Carnap *1937*, 39]

Carnap takes it as a goal of *Logical Syntax* to give an adequate, structural definition of the term consequence in order to provide a logical foundation for mathematics. This is made clear in the following remark: "One of the chief tasks of the logical foundations of mathematics is to set up a formal criterion of validity, that is, to state the necessary and sufficient conditions which a sentence must fulfill in order to be valid (correct, true) in the sense understood in classical mathematics" [Carnap *1937*, 98].

To provide such a definition, he first defines the terms *analytic* and *contradictory*, from which his definition of the proper concept of consequence is defined. Once these two definitions have been given, the definition of consequence runs as follows:

Definition: A sentence θ is a consequence of a set of sentence K if$_{df}$ $K \cup \{\neg\theta\}$ is contradictory.

It is important to note the importance and the specific nature of the terms *analytic* and *contradictory* as they are used in Carnap's early work. Their importance involves the use of these terms in supplying a definition of the proper concept of consequence. As Carnap states:

> We have already seen that the concepts 'demonstrable' and 'refutable' do not fulfill the requirement that they constitute an exhaustive distribution of all logical sentences (which also include mathematical sentences) into mutually exclusive classes. This circumstance provided the reason for the introduction of the concepts 'analytic' and 'contradictory'. [Carnap *1937*, 116].

As regards the specificity of the terms, Carnap equates analytic truth with mathematical truth. For example, as the term is used by Carnap, the Gödel-type formally undecidable sentences are analytic sentences in the language of arithmetic.

In fact, the terms analytic (mathematically true) and logically valid are used interchangeably by Carnap in *Logical Syntax*. This is made evident by remarks Carnap makes throughout *Logical Syntax*. The following is representative of these passages:

> One of the chief tasks of the logical foundations of mathematics is to set up a formal criterion of validity, that is, to state the necessary and sufficient conditions which a sentence must fulfill in order to be valid (correct, true) in the sense understood in classical mathematics. [Carnap *1937*, 98].

The same point also is made in Goldfarb and Ricketts' paper on Carnap: "The analytic sentences turn out to be simply what we would call the logical and mathematical truths. In all essentials Carnap's definitions amount to the same as Tarski's, and Carnap claims that the definitions cannot be formulated within the object languages" [Goldfarb and Ricketts forthcoming, 3]. Here, the authors directly tie Tarski's definitions of the concepts of truth and logical consequence to Carnap's definitions of these terms.[12]

Like Tarski, Carnap also is concerned with the modality condition placed on the relation of logical consequence. In an earlier paper on the same

---

[12]There are important differences between Tarski's work and Carnap's; however, I will here only take up those similarities and differences which aid in the present discussion. A fully detailed discussion of Carnap's work would take the present discussion too far afield.

subject ([Carnap *1935*]), Carnap analyzes the modal terms *impossible* and *necessary* in terms of mathematical definitions. He claims that the modal terms are used to discuss states, events, and conditions, and that they have exact mathematical correlates which are used with reference to the structure of the sentences which describe these states, events, and conditions. This provides a way of "reducing" discussion of the modal notions to a discussion of the form of sentences. The notion of impossibility is correlated with the notion of contradictoriness, and the notion of necessity is correlated with the notion of analyticity. A sentence that is contradictory describes a state that is necessarily false and a sentence that is analytic describes a state that is necessarily true.[13] Thus, as Carnap sees it, the purely mathematical notions of analyticity and contradictoriness provide his account of logical conse-quence and logical validity with the requisite modalities.

7. The Project Defined. We are now in a position to discuss the content of Tarski's project as it is presented in both his 1936 article and in his earlier papers on truth and ω-incompleteness. I will argue that Tarski had a project in mind, when constructing his definitions of logical consequence and logical validity (analyticity), that was decidedly different than the current project. This difference manifests itself, I argue, in the relevant logical definitions being different, as well as in the intended goals of producing these definitions being different.

Much of the modern way of addressing logic comes down to us from the Hilbert school (e.g., the use of purely formal languages). However, in the following quote, Tarski distinguishes his project from Hilbert's:

> But it should be emphasized that the authors mentioned relate this concept not to sentences but to sentential functions with free variables (because in the language of the lower functional calculus which they use there are no sentences in the strict sense of the word) and, connected with this, they use the term "generally valid" instead of the term "correct" or "true". [Tarski *1931*; q.v. *1956*, 199, n. 3][14]

---

[13] I am here leaving off any discussion of what it means for a state to be true or false.

[14] It might be thought that the term 'generally valid' coincides in meaning with the term 'valid' or with the term 'logically valid'; however, as the above discussion has indicated this is not the case. The term 'valid' is used by both Tarski and Carnap synonymously with the terms 'correct' and 'true' in the case where the sentence is from a purely formal language. In fact, the term 'generally valid' never appears in Tarski's 1936 article.

In fact, Tarski had in mind a very specific goal in setting out a deductive science. The professed goal was to arrive at all the truths of a specified theory using only general logical and structurally descriptive concepts (i.e., purely formal concepts).

This goal is clearly stated by Tarski in his paper on ω-incompleteness:

> The formalized concept of consequence will, in extension, never coincide with the ordinary one: the consistency of the system will not prevent the possibility of "structural falsehood." However liberally we interpret the concept of the deductive method, its essential feature has always been (at least hitherto) that in the construction of the system and in particular in the formulation of its rules of inference, use is made exclusively of general logical and structurally descriptive concepts. *If now we wish to regard as the ideal of deductive science the construction of a system in which all true statements (of the given language) and only such are contained*, then this ideal unfortunately cannot be combined with the above view of the deductive method. [Tarski *1933*; *1956*, 295][15]

The italicized portion of the above quotation sums up the motivation for the results that Tarski puts forth in his 1936 paper on logical consequence. For Tarski in [Tarski *1931*] and [Tarski *1936*], as well as for Carnap in [Carnap *1934*], the goal of an "ideal" deductive science was to capture all the truths of a given mathematical theory using purely structural methods, or to capture all the truths of a particular scientific theory using purely structural methods supplemented with non-logical axioms for the science under study. For both Tarski and Carnap, the notion of formality is used synonymously with the notion of structure, and hence they considered logically valid sentences to be sentences which are true based purely on their form, or equivalently, on their structure.

In order to carry out this project, Tarski used an expanded notion of formality (compared to the one in current use), which was reflected in Carnap's writings as well. The notion of formality he used, and along with it the criterion of what it is to be a logical term, was not restricted to the logical constants of the first- or second-order predicate calculus, but rather encompassed the whole of mathematics, conceived broadly as the science of formal languages.[16] The goal of the purely logical deductive sciences was not to deduce the entirety of mathematics from a much smaller logical calculus, but rather to find those sentences and modes of reasoning which could

[15] Italics mine. In his book, *Polish Logic* [Jordan *1989*], Zbigniew Jordan echoes these same sentiments in what amounts to a virtual paraphrasing of Tarski's passage.

[16] Sher's analysis of the notion of formality that is presented in her recent book [Sher *1991*] is similar to the notion that is being explicated here.

be presented in a purely formal manner and which laid the foundation for a particular mathematical discipline. As such, the notion of logical consequence is mapped out in terms of a more general notion of mathematical (again, broadly conceived) inference.

8. Conclusions. I wish to return to Tarski's claim at the end of the 1936 paper: "In the extreme case we could regard all terms of the language as logical. The concept of formal consequence would then coincide with that of material consequence" [Tarski *1936*; *1956*, 419]. We are now in a position to explain the exact nature of this claim. According to the Tarskian account, the claim is true and its validity can be explained by an example Tarski gives in a footnote of his 1936 paper [Tarski *1956*, 419]. Consider the sentence X and the set of sentences $K = \{Y_1, \ldots, Y_n\}$, and allow Z to denote the sentence '$Y_1 \wedge \ldots \wedge Y_n) \rightarrow X$'. Tarski makes the following claims:

[1] The sentence X follows formally from the sentences of the class K if and only if Z is analytical

and

[2] The sentence X follows materially from the sentences of the class K if and only if Z is true

Recall that both Tarski and Carnap consider sentences from a purely formal language (i.e., one that contains no non-logical constants) to be analytic if and only if the sentence is correct or true. Now assume that the sentences X, $Y_1, \ldots, Y_n$ and Z are from a purely formal language (i.e., one that contains no non-logical constants). From these remarks, we arrive at the following equivalence:

[3] The sentence Z is analytical if and only if the sentence Z is true.

And finally, from [1], [2] and [3] we arrive at the desired biconditional:

[F-M] The sentence X follows formally from the sentences of the class K if and only if X follows materially from the sentences of the class K.

The stamp of formality on the inference from K to X is provided by the fact that the sentences X, $Y_1, \ldots, Y_n$ and Z are from a purely formal language, and as such are describable by purely formal (i.e., structural) means (recall the discussion of Carnap). This is the reduction to which Tarski refers in his 1936 paper.

Returning to the examples of the ω-incomplete theory and the Gödel sentences discussed earlier in this paper, an explanation can now be provided for Tarski's analysis of these sentences. Tarski is able to prove in his higher-order system that the universal sentence '$(\forall x)Px$' follows logically from the truth of all of the instantiations, $P(0), \ldots, P(n), \ldots$ for all natural numbers, in that the truth of the relevant sentences is based solely on the structural methods which form the basis of the mathematical theory in which the sentences are formulated, including the size and content of the domain of discourse. He is also able to prove in his system that the Gödel sentence is logically true (analytic) in that it can be shown to be true in the meta-theory on the basis of the purely structural methods that form the foundation of the mathematical theory in which the particular Gödel sentence is formulated.

In contemporary parlance, what Tarski is able to prove is that the relevant sentences are true in the standard model for the formal language in which the sentences are formulated. The stamp of formality, along with the requisite modality, is provided by the fact that all of the terms in the sentences are describable by purely formal (i.e. structural) methods. Of course, this demonstration relies on conflating mathematical and logical expressions in a way which seems odd from a contemporary vantage point, but it is a view that was current during the 1930s in both Tarski's and Carnap's writings.

References.

CARNAP, Rudolf. 1934. *Logische Syntax der Sprache*, Wien, Springer. Translated *as The Logical syntax of language* (A. Smeaton, trans.), London, Routledge & Kegan Paul, 1937.

—. 1935. *Philosophy and logical syntax*, London: Kegan Paul, Trench, Trubner & Company Ltd.

—. 1939. *Foundations of logic and mathematics*, International Encyclopedia of Unified Science, vol. 1, no. 3.

—. 1963. *Intellectual autobiography*, P. A. Schilpp (editor), *The philosophy of Rudolf Carnap,* The Library of Living Philosophers, vol XI (La Salle, Illinois, Open Court), 3–84.

ETCHEMENDY, John. 1990. *The Concept of logical consequence*, Cambridge, Harvard University Press.

FREGE, Gottlob. 1883. *Über den Zweck der Begriffsschrift*, Sitzungsberichte der Jenaischen Gesellschaft für Medicin und Naturwissenschaft für das Jahr 1882, 1–10.

FRIEDMAN, Michael. 1988. *Logical truth and analyticity in Carnap's "Logical Syntax of Language"*, W. Aspray & P. Kitcher (editors), *History and philosophy of modern mathematics* (Minneapolis, University of Minnesota Press), 82–94.

GÖDEL, Kurt. 1931. *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik **38**, 173–198. English translation as *On formally undecidable propositions of Principia Mathematica and related systems* (B. Meltzer, trans., with an Introroduction by R. B. Braithwaite), New York, Basic Books, 1962.

—. 1986. *Collected works*, vol. I: *Publications 1929 – 1936* (Solomon Feferman, John W. Dawson, Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, Jean van Heijenoort, editors), New York, Oxford University Press & Oxford, Clarendon Press.

GOLDFARB, Warren and **RICKETTS**, Thomas. forthcoming. *Carnap and the philosophy of mathematics*, in D. Bell and W. Vossenkuhl (editors), *Science and subjectivity*, Berlin, Akademie Verlag.

HENKIN, Leon. 1950. *Completeness in the theory of types*, The Journal of Symbolic Logic **15**, 81–91.

HODGES, Wilfred. 1986. *Truth in a structure*, Proceedings of the Aristotelian Society **85**, 135–151.

JORDAN, Zbigniew A. 1967. *The development of mathematical logic in Poland between the two wars*, Storrs McCall (editor), *Polish logic 1920 – 1939* (Oxford, Clarendon Press), 346–397.

MCGEE, Vann. 1992. *Review of Etchemendy's "The Concept of Logical Consequence,"* The Journal of Symbolic Logic **57**, 254–255.

—. *Two problems with Tarski's theory of consequence*, Proceedings of the Aristotelian Society **92**, 273–292.

RUSSELL, Bertrand 1903. *The principles of mathematics*, New York, W. W. Norton & Company.

SHER, Gila. 1989. *A concept of Tarskian logic*, Pacific Philosophical Quarterly **70**, 341–368.

—. 1991. *The bounds of logic: A generalized viewpoint*, Cambridge, MA, MIT Press.

—. 1992. "Did Tarski commit Tarski's Fallacy", presented to the Logic Colloqium at University of California at Berkeley.

TARSKI, Alfred. 1931. *The concept of truth in formalized languages*, in [Tarski *1956*] 152–278.

—. 1933. *Some observations on the concepts of ω–consistency and ω – completeness*, in [Tarski *1956*], 279–295.

—. 1936. *On the concept of logical consequence*, in [Tarski *1956*], 409–420.

—. 1956. *Logic, semantics, metamathematics: Papers from 1923 to 1938* (J. H. Woodger, translator), Oxford, Clarendon Press.

—. 1986. *What are logical notions?*, History and Philosophy of Logic **7**, 143–154.

TURQUETTE, Atwell. 1958. *Review of Tarski's "On the Concept of Logical Consequence"*, Philosophical Review **67**, 113-116.

WOLENSKI, Jan. 1989. *Logic and philosophy in the Lvov-Warsaw school*, Dordrecht/Boston, Kluwer Academic Publishers.