Reduction and Tarski's Definition
of Logical Consequence

Jim Edwards

Abstract   In his classic 1936 paper Tarski sought to motivate his definition of
logical consequence by appeal to the inference form: $P(0), P(1), \ldots, P(n), \ldots$
therefore $\forall n\, P(n)$. This is prima facie puzzling because these inferences are
seemingly first-order and Tarski knew that Gödel had shown first-order proof
methods to be complete, and because $\forall n\, P(n)$ is not a logical consequence of
$P(0), P(1), \ldots, P(n), \ldots$ by Taski's proposed definition. An attempt to resolve
the puzzle due to Etchemendy is considered and rejected. A second attempt due
to Gómez-Torrente is accepted as far as it goes, but it is argued that it raises
a further puzzle of its own: it takes the plausibility of Tarski's claim that his
definition captures our common concept of logical consequence to depend upon
our common concept being a reductive conception. A further interpretation of
what Tarski had in mind when he offered the example is proposed, using mate-
rials well known to Tarski at the time. It is argued that this interpretation makes
the motivating example independent of reductive definitions which take natural
numbers to be higher-order set theoretic entities, and it also explains why he did
not regard the distinction between defined and primitive terms as pressing, as
was the distinction between logical and nonlogical terms.

1   Introduction

In his classic 1936 paper [12] Tarski sought to motivate acceptance of his definition
of logical consequence by discussion of an example. He wrote,

> Some years ago I gave a quite elementary example of a theory which shows
> the following peculiarity: among its theorems there occur such sentences as:
>
> A_0.   0 *possess the given property P*,
> A_1.   1 *possess the given property P*,
>
> and, in general, all particular sentences of the form

> A_n.   *n possess the given property* P,

where 'n' stands for any symbol which denotes a natural number in a given (e.g., decimal) number system. On the other hand, the universal sentence,

> A.   *Every natural number possesses the given property* P,

cannot be proved on the basis of the theory in question by means of the normal rules of inference. This fact seems to me to speak for itself. It shows that the formalized [i.e., proof-theoretic] 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 [i.e., pretheoretic] sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true. ([12], pp. 410–11)

As Tarski in effect goes on to note, the above fact "speaks for itself" only in the light of Gödel's incompleteness result. For we might conjecture that we could close the proof-theoretical gap by adding a further rule of inference, what Tarski called "the rule of infinite induction according to which the sentence A can be regarded as proved provided all the sentences $A_0, A_1, \ldots, A_n, \ldots$ have been proved" ([12], p. 411). He summarily dismissed the rule of infinite induction on the grounds that it is objectionably infinitistic—its application would require an infinite set of sub-proofs to provide the requisite premises.[1] Tarski then considered adding to the base theory T an infinite series of finitistic proof rules $R_0, R_1, \ldots, R_i, \ldots$ such that the premise of R_0 is the arithmetization of the provability of $A_0, A_1, \ldots, A_n, \ldots$ in the theory T, and the conclusion is A, and the premise of R_1 is the arithmetization of the provability of $A_0, A_1, \ldots, A_n, \ldots$ in the theory $\{T \cup R_0\}$, and the conclusion is A, and the premise of R_2 is the arithmetization of the provability of $A_0, A_1, \ldots, A_n, \ldots$ in the theory $\{T \cup R_0 \cup R_1\}$, and the conclusion is A—and so on through R_3, R_4, etc. Tarski invoked Gödel to refute the conjecture that such procedures can close all the proof-theoretical gaps.

> By making use of the results of K. Gödel we can show that this conjecture is untenable. In every deductive theory (apart from certain theories of a particularly elementary nature), however much we supplement the ordinary 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 the theory on the basis of the accepted rules of inference. ([12], pp. 412–13)

Tarski then proceeds to introduce his own model-theoretic definition of logical consequence:

> The *sentence X* follows logically *from the sentences of the class* K *if and only if every model of* [all the sentences in] *the class K is also a model of the sentence X*.[2] ([12], p. 417, Tarski's italics)

Prima facie, Tarski's dialectic is puzzling. It looks as if he thought

(i) "[I]ntuitively it seems certain that the universal sentence A follows in the usual [i.e., pretheoretic] sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true."

(ii) Gödel has shown that proof-rules cannot capture all such inferences.

Hence we are motivated to accept his model-theoretic conception of logical consequence. So we seem entitled to expect

(iii) Such inferences come out valid by Tarski's model theoretic definition of logical consequence.

This is puzzling for two reasons:

(iv) The inference A_0, A_1, etc. therefore A is *not* valid according to Tarski's proposed definition. If we standardly take '0', '1', '2', etc. as proper names and therefore as nonlogical constants it is easy to find models in which $A_0, A_1, \ldots, A_n, \ldots$ are all true but A false. Consider, for example, a model M of A_0, A_1, etc. and A whose domain is the natural numbers, where 'natural number' has its usual interpretation, but '0', '1', '2', etc. are all assigned 0, and the extension assigned 'P' is $\{0\}$. Obviously, A_0, A_1, etc. are all true-in-M, and A is not true-in-M.

(v) Such examples are seemingly first-order. But as Tarksi knew, Gödel had shown that proof-theoretic consequence and model-theoretic consequence coincide for first-order structures.

In this paper we will resolve these puzzles, along with a subsequent puzzle which arises from the solution to the first puzzles.

2 Two Attempts to Resolve These Puzzles

I will consider two attempts to resolve these puzzles: one due to Etchemendy ([1], [2]) and the other due to Gómez-Torrente [4]. I shall reject Etchemendy's "solution" as not consistent with Tarski's text and clear intentions. I shall accept Gómez-Torrente's solution as far as it goes, but I shall argue that it raises a further puzzle of its own: it takes the plausibility of Tarski's definition of logical consequence to presuppose logicist reductions of arithmetic terms to logical terms, which restricts the scope of Tarski's definition. Etchemendy's and Gómez-Torrente's respective "solutions" are polar opposites; my own will occupy a place between them.

First we consider Etchemendy. He notes that the inference from A_0, A_1, etc. to A

> comes out valid, model-theoretically, if we include in [the set of logical constants] the expression "every natural number" as well as the collection of numerals '0', '1', '2', ... I assume this is why Tarski does not consider his account subject to precisely the same criticism he directs at the [proof-theoretic] definition. ([2], p. 85)

Of course, if 'every natural number' and '0', '1', '2', etc. are logical constants then A is indeed a Tarskian logical consequence of A_0, A_1, etc. However, this solution would lead to disaster for Tarski's project. Etchmendy notes elsewhere:

> Gödel sentences are a bit trickier, due to their potential variety: all that we can really say is that they will indeed come out as consequences of their corresponding theories if we treat all expressions in the language as logical constants. Unfortunately, this involves a certain trivialization of Tarski's Analysis. For with this choice of logical constants, a true sentence is a logical consequence of any set of sentences whatsoever. ([1], p. 73)

Actually, it's a complete trivialization, in Tarski's own view. If all the primitive terms of the language are counted as logical constants, then, so Tarski thought, Tarskian logical consequence collapses into material consequence, something he pointed out himself in the same paper in which he offered his definition of logical consequence:

> In the extreme case we could regard all the terms of the language as logical [i.e., as logical constants]. The concept of *formal* [= logical] consequence would then coincide with that of *material* consequence. The sentence X

would in this case follow from the class K of sentences if either X were true or at least one sentence of the class K were false.[3] ([12], p. 419, Tarski's italics)

Tarski closes the paper by remarking that the distinction between logical and non-logical constants is the next big unsolved problem. If Etchemendy were right it would have been solved. Gödelian ω-inferences require that there are no nonlogical constants!—well, except perhaps in those deductive theories "of a particularly elementary nature" to which Gödel's result does not apply.

I regard it as incredible that Tarski, in the same paper, should have raised the issue of inferences from A_0, A_1, etc. to A, used it to dismiss proof-theoretic conceptions of logical consequence, proposed his model-theoretic alternative, and not thought that it could meet the challenge. But I regard it as equally incredible that he should have thought that his model-theoretic conception met the challenge in a way that, in his view, collapsed logical consequence into material consequence, especially as, in that same paper, he is explicitly aware of the danger, and he explicitly regards the division between logical and nonlogical constants as an open question.

As Etchemendy reads Tarski, 'natural number' and '0', '1', '2', etc. are all logical constants. As Gómez-Torrente reads Tarski, none are.

> The solution suggested by the textual evidence is that when he gives his motivating example Tarski is not thinking of the arithmetical expressions as primitives, but as defined terms; defined, that is, with the help of logical constants, within the framework of a sufficiently powerful logical theory. ([4], p. 136)

Gómez-Torrente takes his cue from the first sentence of the passage from Tarski quoted above: "Some years ago I gave a quite elementary example of a theory which shows the following peculiarity." Tarski was there referring to his [13]. In the formal language discussed in [13], 'natural number', '0', '1', '2', etc. do not appear. The language in which the theory is expressed contained the primitive sentential functions negation and material implication, the universal quantifier, and nothing else except variables. The variables were sorted into types. Thus 'x_1^1', 'x_2^1', 'x_3^1', etc. were all first-order variables whose values were individuals from the domain. And 'x_1^2', 'x_2^2', 'x_3^2', etc. were all second-order variables whose values were sets of individuals from the domain. And so on for variables of all finite orders. Any of these variables could be quantified over. An atomic open sentence took the form $x_i^{n+1}(x_j^n)$. In addition to the primitive signs '$\rightarrow$', '$\neg$', and '$\forall$', he introduced as defined signs '$\exists$', '$\vee$', '&', '$\leftrightarrow$', and '$=$'. The theory defined on this language was a set of standard axioms—propositional and quantification axioms, plus axioms of comprehension and extensionality for each type, and an axiom of infinity for the objects of lowest type—and the consequences of these axioms under substitution, detachment, universal introduction, and elimination. The terms 'natural number', '0', '1', '2', etc. did not appear in this language as primitives, though they may be introduced as defined terms. So, as Gómez-Torrente reads Tarski, 'natural number', '0', '1', '2', etc. in [12] are *defined* terms. As such they are not, pace Etchmendy, logical constants, but are to be eliminated from $A_0, A_1, \ldots, A_n, \ldots$ and A before Tarski's definition is applied to see whether A is a logical consequence of A_0, A_1, etc. Tarski draws our attention to "the necessity of eliminating any defined signs which may possibly occur in the sentences concerned, i.e., of replacing them by primitive signs" ([12], p. 415). Once 'natural number', '0', '1', '2', etc. have been eliminated from A

and A_0, A_1, etc., Gómez-Torrente argues, the puzzles (iv) and (v) above may be resolved. Puzzle (iv) is resolved. The numerals '0', '1', etc. and the predicate 'natural number' are not primitive signs and so have been eliminated in favor of basic vocabulary before the definition of logical consequence is applied. So, "since the only extra-logical constants subject to reinterpretation in the test for logical consequence will appear, if there are any, in the predicate P" ([4], p. 136), A will be a logical consequence of A_0, A_1, etc. by Tarski's definition. Puzzle (v) is also resolved. A and A_0, A_1, etc., once primitives have been eliminated, are revealed as higher-order; to them Gödel's completeness result does not apply.

However, it is worth noting, to avoid exegetical confusion, that the relation between Tarski's [12] and his [13] is more complex than has been revealed so far. In his [13], Tarski drew *particular* consequences from the theory described and he showed, in the light of those consequences, that the theory is ω-consistent but not ω-complete. But these particular consequences are *not* the example Gómez-Torrente takes Tarski to have in mind as A and A_0, A_1, etc. in his [12].[4] The particular consequences discussed in [13] take the following forms:[5]

B_0 $\quad \forall x_1^2 (\forall x_1^1 \neg x_1^2(x_1^1) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1))$

B_1 $\quad \forall x_1^2 (\exists x_1^1 \forall x_2^1 (x_1^2(x_2^1) \rightarrow x_2^1 = x_1^1) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1))$

$\ldots$

B_n $\quad \forall x_1^2 (\exists x_1^1 \ldots \exists x_n^1 \forall x_{n+1}^1 (x_1^2(x_{n+1}^1) \rightarrow$

$$(x_{n+1}^1 = x_1^1 \vee \cdots \vee x_{n+1}^1 = x_n^1)) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1))$$

$\ldots .$

Tarski [13] showed that whereas each of the above is a proof-theoretic consequence of the axioms, the following is not:

B $\quad \forall x_1^2 (\forall x_1^3 (\forall x_2^2 \forall x_3^2 ((\forall x_1^1 \neg x_2^2(x_1^1) \vee (x_1^3(x_2^2) \, \& \, \exists x_1^1 \forall x_2^1 (x_2^2(x_2^1) \leftrightarrow$

$$(x_3^2(x_2^1) \vee x_1^1 = x_2^1)))) \rightarrow x_1^3(x_2^2)) \rightarrow x_1^3(x_1^2)) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1)).$$

B and B_0, B_1, ..., B_n, ... would not have served Tarski's [12] purpose since they are not examples of A and A_0, A_1, ..., A_n, ..., even when '0', '1', etc. and 'natural number' have been eliminated by definition from the latter. For B_0, B_1, ..., B_n, ... did not exhibit a common logical form. Rather, a string of first-order existential quantifiers grew by one each time as the series progressed, and the matrix which they governed changed in tandem to provide a further argument place. In effect B_0 stated of the empty lowest-order set, that there is something in the domain which is not a member of it. B_1 stated of each lowest-order set with at most one member, that there is something in the domain which is not a member of it. And B_n stated of each lowest-order set with at most n members, that there is something in the domain which is not a member of it. The supposed analogue of the conclusion A would have been of quite a different form again, involving a third-order quantifier.

B $\quad \forall x_1^2 (\forall x_1^3 (\forall x_2^2 \forall x_3^2 ((\forall x_1^1 \neg x_2^2(x_1^1) \vee (x_1^3(x_2^2) \, \& \, \exists x_1^1 \forall x_2^1 (x_2^2(x_2^1) \leftrightarrow$

$$(x_3^2(x_2^1) \vee x_1^1 = x_2^1)))) \rightarrow x_1^3(x_2^2)) \rightarrow x_1^3(x_1^2)) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1)).$$

The content of B is hard to read straight off. Its structure can be simplified to

$$\forall x_1^2 (\forall x_1^3 \varphi(x_1^2, x_1^3) \rightarrow \exists x_1^1 \neg x_1^2(x_1^1))$$

where the condition $\forall x_1^3 \varphi(x_1^2, x_1^3)$ ensures that x_1^2 has at most a finite number of members. Thus B states in effect that no finite lowest-order set contains all the members of the domain. Tarski shows that, relative to his set of logical axioms and rules of inference, B_0, B_1, etc. are all proof-theoretic theorems, but B is not. However, B and B_0, B_1, etc. are not of the forms A and A_0, A_1, etc. even after we have eliminated '0', '1', etc. and 'natural number' from A and A_0, A_1, etc. by standard definitions. So we need to set B and B_0, B_1, etc. aside as a red herring, from the point of view of [12], as understood by Gómez-Torrente.

3 A Further Puzzle

Puzzlement resolved? Not really. As Gómez-Torrente acknowledges, the examples he describes carry a commitment to the reductive definitions of numeric terms favored by logicists.

> If the predicate 'to be a natural number' and the numerals '0', '1', '2', etc. are defined in the logicist fashion within the framework of an appropriate logical theory, A will follow from A_0, A_1, etc. according to Tarski's definition . . . . ([4], p. 136)

Those definitions identify numbers as higher-order set theoretical entities—for example, $0 = \{x : x \neq x\}$, $1 = \{0\}$, $2 = \{0, 1\}$, etc. A number of such definitions were current at the time. If such were the *only* examples Tarski had in mind, then he would have been committed to holding that our intuition that A follows from A_0, A_1, etc. is accounted for by his definition *only if we accept that numbers are higher-order set-theoretic entities*. Certainly such was the view of many, including Tarski himself, at the time—he spoke of such reductions being "one of the grandest achievements of recent logical investigations" (Tarski [9], p. 81). However, it would be surprising if Tarski thought his definition of logical consequence itself presupposed such a reduction, as distinct from being merely consistent with them. After all, Tarski expected that Gödel would agree that, intuitively, A follows from A_0, A_1, etc. And Tarski hoped that Gödel too could accept his definition of logical consequence as capturing such intuitions. But Tarski also recognized that numerals are primitives in the language of Gödel's [3] and in the deductive system P which Gödel expressed using that language. He wrote of the formal language of [3], comparing it to that of his own [11]:

> Apart from certain differences of a "calligraphical" nature, the only distinction lies in the fact that in the system P, in addition to the logical constants, certain constants belonging to the arithmetic of the natural numbers also occur. ([11], p. 247–48, ft. 1)

Since the only arithmetic constants in the language on which P is defined are '0' and 's', the successor function, Tarski himself was reading these as primitive symbols, and as *non*logical constants of the language in which theory P is written. I shall set out system P shortly, but it is easy to see—from axioms I1 – I3, and rule of inference R1 below—that the natural numbers, 0 and all its successors, are the denizens of the *lowest-order* domain and so cannot be defined in that language in a logicist fashion as higher-order entities. Tarski certainly thought his definition of logical consequence applicable to the sentences of Gödel's formal language—for if not it would be an obvious and grave defect. And it is reasonable to suppose that Tarski thought that

logicians who took Gödel's formal language as ontologically basic and who read A and A_0, A_1, etc. in terms of that language, would have agreed that A intuitively follows from A_0, A_1, etc. and would have expected this intuition to be honored by a satisfactory definition of logical consequence. We have not yet seen how to do this. This is the further puzzle.

4 The Last Puzzle Resolved

To resolve this puzzle, I shall argue that when drawing our attention to A and A_0, A_1, etc. Tarski had examples in mind other than those described by Gómez-Torrente. These other examples retain at least some arithmetic terms as primitive. We require that these additional examples show the following features:

1. The examples are higher-order once defined terms in A and A_0, A_1, etc. are replaced by primitive vocabulary, thus avoiding Gödel's completeness result for first-order theories.[6]
2. However, some arithmetic terms appear as primitives, and natural numbers are the urelements of the lowest-order domain.
3. It is intuitive that the counterpart of A (i.e., A once defined terms have been eliminated) is a logical consequence of the counterparts of A_0, A_1, etc. in these examples.
4. The counterpart of A is a logical consequence of the counterparts of A_0, A_1, etc. according to Tarski's own definition of logical consequence, given a plausible distinction between logical and nonlogical constants.
5. The counterparts of A_0, A_1, etc. are all proof-theoretic consequences of some presupposed theory, but A is not.
6. Gödel's result can be used to show that proof-theoretic methods cannot capture all such inferences.

I shall develop an example satisfying (1) to (6), an example which Tarski might plausibly also have had in mind since it draws on features well known to him at the time.

Of course Gödel himself provided an example close to A and A_0, A_1, etc., an example employing, according to Tarski, '0' and 's' as arithmetic primitives and of the following form:

C_0 $F(0)$

C_1 $F(s0)$

C_2 $F(ss0)$

etc.

C $\forall x F(x)$

Gödel showed that each of C_0, C_1, etc. is a proof-theoretic consequence of P but C is not. But this is not the example we are looking for. Because the quantifier of C is unrestricted, condition (4) is not met: C is not a logical consequence of C_0, C_1, etc. according to Tarski's own semantic conception of logical consequence.[7] As Gómez-Torrente notes:

> If the arithmetical expressions are not logical constants when they are primitives of our formalization of arithmetic, C will not be declared a logical consequence of the set of sentences C_0, C_1, etc. by Tarski's definition. ([4], p. 136, transposed to my notation)

However, it is easy to construct an example satisfying (1) to (6) using Gödel's C and C_0, C_1, etc. as a template. We merely need to restrict the quantifier in C to the natural numbers. The language in which Gödel framed P, C and C_0, C_1, etc. was higher-order with primitive open sentences of the form $y^{n+1}(x^n)$. The language contained just five primitive constants: '$\neg$', '$\vee$', '$\forall$', '0', and 's'. Gödel availed himself of the defined signs: '&', '$\rightarrow$', '$\leftrightarrow$', '$\exists$', '='. In this higher-order language we can define 'natural number' as follows.

Definition 4.1

(Df) y^1 is a natural number $=$ df $\forall x^2((x^2(0) \,\&\, \forall x^1(x^2(x^1) \rightarrow x^2(sx^1))) \rightarrow x^2(y^1))$.

Df ensures that y^1 is a member of the smallest set which contains 0 and contains the successor of any member. Subject to the comments in the next paragraph, that just is the set which contains 0, its successor $s0$, its successor $ss0$, etc. and nothing else. This is the set of natural numbers.

   Df ensures that y^1 is a member of the smallest set which contains the natural numbers in the domain of '$\forall x^2$'. To ensure that the set contains nothing else besides the natural numbers we require that the domain of '$\forall x^2$' be the powerset of the domain of '$\forall x^1$'. Otherwise the smallest set which contains 0 and all its successors may contain other entities too—there being no smaller set in the domain of '$\forall x^2$'. In so-called Henkin models (which [7] calls "general models") the domain of '$\forall x^2$' is required to be only *some* set of subsets of the domain of '$\forall x^1$'—and in general the domain of '$\forall x^i$' is *some* set of subsets of the domain of '$\forall x^{i-1}$'. In Henkin's semantics we cannot suppose that, in all models, the smallest set containing the bearers of '0', '$s0$', '$ss0$', etc. contains nothing else besides $0, s0, ss0$, etc. By contrast, in so-called full models, the domain of '$\forall x^2$' is the "full" powerset of the domain of '$\forall x^1$' and in general the domain of '$\forall x^i$' is the powerset of the domain of '$\forall x^{i-1}$'. Hence for Df to define the natural numbers we require that the semantics of the language be given in terms of full models, not in terms of the more general notion of Henkin models.

   Tarski did not have the distinction between Henkin models and full models to hand in [12]. So a question arises regarding his proposed definition of logical consequence:

> The sentence X is a logical consequence of the sentences of the class K if and only if every model in which all the sentences of the class K are true is also a model in which X is true.

Can we properly attribute to him a semantics of full models? We can, for two reasons. Firstly, he commits himself to a semantics of full models for the higher-order language of his [13]. He defines '$x_j^i = x_k^i$' as $\forall x_1^{i+1}(x_1^{i+1}(x_j^i) \rightarrow x_1^{i+1}(x_k^i))$. For this definition to assign the identity function to '=' requires that the domain of '$\forall x^{i+1}$' be the powerset of the domain providing the values of 'x_j^i' and 'x_k^i'. Secondly, he commits himself to a semantics of full models when he takes Gödel to have shown that Gödel's theory P is proof-theoretically incomplete. P is a higher-order system with second-order Peano axioms. And Henkin showed that such systems are proof-theoretically *complete* with respect to Tarski's definition of logical consequence and a semantics of Henkin models [6]. If we were to read Tarski's definition of logical consequence in terms of Henkin models, there would be no examples where counterparts of A, A_0, A_1, etc. are higher-order, where it is *not* the case that A_0, A_1, etc.

$\vdash_P A$, and yet A *is* a logical consequence of A_0, A_1, etc. on Tarski's proposed defini-tion.[8] But clearly Tarski thinks there are such examples. So we should read Tarski's proposed model-theoretic definition of logical consequence as referring to full mod-els, not to Henkin models, when applied to higher-order languages. Hence we can take the semantics of the language upon which P is defined to be a semantics of full models, and we can take Df to be a definition of natural number.

We can now at last give an example which satisfies conditions (1) to (6) above. Taking Gödel's C and C_0, C_1, etc. as our template, consider the following related sentences:

D_0 $F(0)$

D_1 $F(s0)$

D_2 $F(ss0)$

etc.

and

D $\forall y^1(\forall x^2((x^2(0)\ \&\ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1)) \to F(y^1))$.

Conditions (1) to (6) are met by D and D_0, D_1, D_2, etc. By inspection, (1) to (4) are satisfied.

1. D and D_0, D_1, etc. are higher-order once defined terms in A and A_0, A_1, etc. are replaced by primitive vocabulary.

2. '0' and 's' are arithmetic primitives in the language of D and D_0, D_1, etc., and natural numbers are the urelements of the lowest-order domain.

3. It is intuitive that D is a logical consequence of D_0, D_1, etc. in these exam-ples. We can gloss D informally as: 'For all natural numbers y^1, $F(y^1)$'— which is of the form of Tarski's A. And we can gloss D_0, D_1, etc. informally as '$F(0)$', '$F(1)$', etc., which are of the form of Tarski's A_0, A_1, etc.

4. D is a Tarskian logical consequence of D_0, D_1, etc.—taking models to be full models. We can see this because the only nonlogical constants are '0' and 's', and the antecedent of D selects the elements of the domain which belong to the smallest set containing the bearer of '0', the value of '$s0$', the value of '$ss0$', etc. Thus every full model in which the sentences D_0, D_1, etc. are all true is a model in which D is true also. Thus D is a logical consequence, by Tarski's definition, when we read 'model' as full model, of D_0, D_1, etc.

It remains to show the following:

5. D_0, D_1, etc. are all proof-theoretic consequences of some presupposed the-ory, but D is not.

6. Gödel's result can be used to show that proof-theoretic methods cannot cap-ture all such inferences.

We can show that (5) is satisfied relative to Gödel's theory P. The axioms of P are given as open sentences or schemas:

I   1 $\neg(sx^1 = 0)$

    2    $(sx^1 = sy^1) \to (x^1 = y^1)$

    3    $(x^2(0)\ \&\ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to \forall x^1(x^2(x^1))$

II   1 $p \lor p \to p$

    2    $p \to p \lor q$

    3    $p \lor q \to q \lor p$

    4    $(p \to q) \to (r \lor p \to r \lor q)$

III   1 $\forall x^n(A) \to A(c/x^n)$

       where 'c' is any sign of the same type as 'x^n' and 'c' does not contain any variable that is bound in A at a place where 'x^n' is free.

    2    $\forall x^n(B \lor A) \to B \lor \forall x^n A(x^n)$ provided 'x^n' is not free in B.

IV $\exists x^{n+1}(\forall x^n(x^{n+1}(x^n) \leftrightarrow A))$ provided 'x^{n+1}' does not occur free in A.

V $\forall x^n(x^{n+1}(x^n) \leftrightarrow y^{n+1}(x^n)) \to x^{n+1} = y^{n+1}$.

The rules of inference of P are:

R1   From A, we may infer $\forall v A$, where v is any variable of any order.

R2   From $\neg A \lor B$ and A we may infer B.

Given that 'F' is Gödel's predicate, we have immediately that D_0, D_1, etc. are proof-theoretic consequences of P since they are identical to C_0, C_1, etc. which Gödel has shown to be proof-theoretic consequences of P. It remains to show that D is not a proof-theoretic consequence of P.

Proof   Suppose for reductio that D is a proof-theoretic consequence of P. We have as theorems of P:

1. $\forall y^1(\forall x^2((x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1)) \to F(y^1))$

                              i.e., D

2. $\forall x^1(x^2(x^1)) \to x^2(y^1)$ From axiom III1

3. $((x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to \forall x^1(x^2(x^1))) \to$
                 $((x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1))$

               From (2), axiom II4, and def. of '$\to$', by R2

4. $(x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1)$

               From (3), axiom I3, and def. of '$\to$', by R2

5. $\forall x^2((x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1))$

               From (4), by R1

6. $\forall x^2((x^2(0) \ \& \ \forall x^1(x^2(x^1) \to x^2(sx^1))) \to x^2(y^1)) \to F(y^1)$

               From (1), axiom III1, and def. of '$\to$', by R2

7. $F(y^1)$ From (5), (6), and def. of '$\to$', by R2

8. $\forall y^1 F(y^1)$ From (7), by R1

But Gödel has proved that (8) is *not* a proof-theoretic consequence of P, if P is ω-consistent. Therefore D, that is, (1), is not a proof-theoretic consequence of P, if P is ω-consistent.   □

Since D_0, D_1, etc. are proof-theoretic consequences of P but D is not, if P is ω-consistent, condition (5) is satisfied.

It follows that condition (6) is also satisfied: the example generalizes to block a proof-theoretic account of logical consequence. We can rewrite D more fully as

$$D \quad \forall y^1(\forall x^2((x^2(0) \,\&\, \forall x^1(x^2(x^1) \rightarrow x^2(sx^1))) \rightarrow x^2(y^1)) \rightarrow \neg\mathrm{Prov}_P(y^1, m))$$

where 'm' codes for '$\forall y^1 \neg\mathrm{Prov}_P(y^1, m)$'. Gödel has shown that whatever recursively specifiable set of axioms we add to P to form an extended system P^*, we can construct an object language sentence '$\forall y^1 \neg\mathrm{Prov}_{P*}(y^1, m^*)$', where '$m^*$' codes for this sentence, which is not a proof-theoretic consequence of P^*, if P^* is consistent. From this we can construct an example:

$D_0^* \quad \neg\mathrm{Prov}_{P*}(0, m^*)$

$D_1^* \quad \neg\mathrm{Prov}_{P*}(s0, m^*)$

$D_2^* \quad \neg\mathrm{Prov}_{P*}(ss0, m^*)$

etc.

and

$$D^* \quad \forall y^1(\forall x^2((x^2(0) \,\&\, \forall x^1(x^2(x^1) \rightarrow x^2(sx^1))) \rightarrow x^2(y^1)) \rightarrow \neg\mathrm{Prov}_{P*}(y^1, m^*)).$$

The arguments to show that D and D_0, D_1, etc. satisfy conditions (1) to (5) apply to D^* and D_0^*, D_1^*, etc. Hence our final condition (6) is also satisfied: the example D and D_0, D_1, etc. generalizes to block a proof-theoretic account of logical consequence.[9]

5 Conclusion

The example D and D_0, D_1, etc. and others of the ilk D^* and D_0^*, D_1^*, etc. are all constructed from material Tarski was thoroughly familiar with, and it would have been clear to him that they satisfied conditions (1) to (6). So charity requires that in drawing our attention to A and A_0, A_1, etc. he had such examples in mind as well as those described by Gómez-Torrente. Charity requires this because the motivation he offers for his definition of logical consequence is then independent of any personal commitment to identifying numbers as higher-order set-theoretical objects. It thus complements Gómez-Torrente's account by explaining why Tarski could claim to be explicating the concept of logical consequence common to mathematical logicians—whether or not they made such identifications. Charity requires this also because it explains a further feature of Tarski's [12]. The application of Tarski's definition of logical consequence to a sentence X and a set of sentences K depends upon two parameters: which terms, if any, in X and the members of K are defined terms as against primitive terms, and, having eliminated defined terms, which, if any, of the remaining primitives are logical constants as against nonlogical constants? Each parameter may affect the output of the definition. Thus, taking '0', '1', etc. and 'natural number' as primitives but nonlogical constants, A is not a Tarskian logical consequence of A_0, A_1, etc., but taking them as logical constants (Etchemendy) or as defined terms (Gómez-Torrente), A is a Tarskian logical consequence of A_0, A_1, etc. Tarski was exercised by the need to determine more precisely the division between logical and nonlogical constants. He flagged the topic up in [12] as the next big problem and returned to it in his posthumously published [14]. But he does not seem to have felt that the division between primitive and defined terms was similarly urgent. He allowed that in some languages '0', '1', etc. were taken as primitive and in others as defined. Perhaps that was because, as far as the crucial inferences of the

grammatical if not logical form A and A_0, A_1, etc. were concerned, it didn't matter to the output of his definition whether '0', '1', etc. (or '0', 's') were taken as primitive or defined, so long as 'natural number' was taken as a defined term.

Notes

1. Interestingly, Tarski was aware that if the axioms of T are the Peano axioms, then T and the rule of infinite induction is proof-theoretically complete: "In the case of certain elementary deductive sciences, [the enlargement of the theory by the addition of the rule of infinite induction] is so great that the class of theorems becomes a complete system and coincides with the class of true sentences. Elementary number theory provides an example, namely, the science in which all variables represent names of natural or whole numbers and the constants are the signs from the sentential and predicate calculi, the signs of zero, one, equality, sum, product, and possibly other signs defined with their help" ([11], pp. 260–61).

2. Tarski conceived of a model as mathematicians do: a model of a set of sentences K is a structure such that all members of K receive the value true. Nowadays logicians standardly conceive of a model of a set of sentences K as a structure providing a domain and interpreting the members of K in such a way that, in general, a member of K may receive the value true or alternatively the value false. Hence the standard formulation of Tarskian logical consequence has become: The sentence X is a logical consequence of the sentences of the class K if and only if every model in which all the sentences of the class K are true is also a model in which X is true. There is also an issue of whether in [12] Tarski considered allowing the domains of models to vary, as he did in his [10]. This issue will not concern us here.

3. Logical consequence collapses into material consequence only if Tarski does not allow the domain to vary when we regard all the terms of the language as logical. This is another controversial point in exegesis of Tarski's paper. However, it is not one we are concerned with here. It is well discussed in Hodges [8] and Gómez-Torrente [4].

4. I thank a referee for making this point clear to me.

5. I've transcribed Tarski's notation, and will later Gödel's, into something more familiar.

6. Tarski wrote in a passage quoted above:
   > In every deductive theory (apart from certain theories of a particularly elementary nature) . . . ([12], pp. 412–13).
   It is reasonable to suppose that by "certain theories of a particularly elementary nature" he meant first-order theories.

7. Gödel did not need to restrict the quantifier in C because, as we shall shortly see, his theory P incorporates second-order Peano axioms. Such a theory is categorical: all models are isomorphic to the natural numbers.

8. On the intended interpretation of the language upon which P is defined, C codes for the metalinguistic statement 'C is not provable in P'. Clearly, in Henkin models in which C_0, C_1, etc. are all true but C false, C is not equivalent to that metalinguistic statement. In some Henkin models, extra elements in the smallest set containing the

natural numbers prevent the object language arithmetic function 'Prov$_P(x, m)$'—where m codes for '$\forall y^1 \neg$Prov$_P(y^1, m)$'—coding for the provability in P of C.

9. We could generate other examples somewhat closer to A and A_0, A_1, etc. In the passage first quoted Tarski speaks of 'n' in A_n being 'any symbol' in any numeral system. We could take '0' to '9' to be the nonlogical primitives, together with a primitive multiadic function symbol taking us from, for example, $\langle$'1', '0', '7'$\rangle$ to the number 107. We would have to bring the definition of 'natural number' into line with this number system to complete our example.

References

[1] Etchemendy, J., "Tarski on truth and logical consequence," *The Journal of Symbolic Logic*, vol. 53 (1988), pp. 51–79. Zbl 0645.03002. MR 89g:01051. 51

[2] Etchemendy, J., *The Concept of Logical Consequence*, Harvard University Press, Cambridge, 1990. Zbl 0743.03002. MR 92f:03003. 51

[3] Gödel, K., "On formally undecidable propositions of *Principia Mathematica* and related systems," pp. 596–616 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 54

[4] Gómez-Torrente, M., "Tarski on logical consequence," *Notre Dame Journal of Formal Logic*, vol. 37 (1996), pp. 125–51. Zbl 0855.03004. MR 97c:03004. 51, 52, 53, 54, 55, 60

[5] Henkin, L., "Completeness in the theory of types," pp. 51–63 in *The Philosophy of Mathematics*, edited by J. Hintikka, Oxford Readings in Philosophy, Oxford University Press, London, 1969. Zbl 0181.29501. MR 40:2517. 61

[6] Henkin, L., "Completeness in the theory of types," *The Journal of Symbolic Logic*, vol. 15 (1950), pp. 81–91. Reprinted in [5]. Zbl 0039.00801. MR 12,70b. 56

[7] Henkin, L., "The discovery of my completeness proofs," *The Bulletin of Symbolic Logic*, vol. 2 (1996), pp. 127–58. Zbl 0882.03001. MR 97c:03005. 56

[8] Hodges, W., "Truth in a structure," *Proceedings of the Aristotelian Society* (1986), pp. 135–51. 60

[9] Tarski, A., *Introduction to Logic and to the Methodology of Deductive Sciences*, Oxford University Press, New York, 1941. Zbl 0025.00401. MR 2,209a. 54

[10] Tarski, A., in collaboration with A. Mostowski and R. Robinson, "A general method in proofs of undecidability," pp. 1–35 in *Undecidable Theories*, 2d printing, Studies in Logic and the Foundations of Mathematics, North-Holland Publishing Co., Amsterdam, 1968. Zbl 0187.27702. MR 15,384h. 60

[11] Tarski, A., "The concept of truth in formalized languages," pp. 152–278 in *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, edited by J. Corcoran, Hackett Publishing Co., Indianapolis, 2d edition, 1983. Translated by J. H. Woodger. MR 85e:01065. 54, 60

[12] Tarski, A., "On the concept of logical consequence," pp. 409–20 in *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, edited by J. Corcoran, Hackett Publishing Co., Indianapolis, 2d edition, 1983. Translated by J. H. Woodger. MR 85e:01065.   49, 50, 52, 53, 54, 56, 59, 60

[13] Tarski, A., "Some observations on the concepts of ω-consistency and ω-completeness," pp. 279–95 in *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, edited by J. Corcoran, Hackett Publishing Co., Indianapolis, 2d edition, 1983. Translated by J. H. Woodger. MR 85e:01065.   52, 53, 56

[14] Tarski, A., "What are logical notions?" *History and Philosophy of Logic*, vol. 7 (1986), pp. 143–54. Edited by J. Corcoran. Zbl 0622.03004. MR 88b:03010.   59

Acknowledgments

Department of Philosophy
University of Glasgow
Glasgow G12 9AS
UNITED KINGDOM
J.Edwards@Philosophy.arts.gla.ac.uk