THE REVIEW OF SYMBOLIC LOGIC
Volume 2, Number 2, June 2009

REREADING TARSKI ON LOGICAL CONSEQUENCE

MARIO GÓMEZ-TORRENTE
Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México

Abstract. I argue that recent defenses of the view that in 1936 Tarski required all interpretations of a language to share one same domain of quantification are based on misinterpretations of Tarski's texts. In particular, I rebut some criticisms of my earlier attack on the fixed-domain exegesis and I offer a more detailed report of the textual evidence on the issue than in my earlier work. I also offer new considerations on subsisting issues of interpretation concerning Tarski's views on the logical correctness of certain omega-arguments, on the Tarskian proof that Etchemendy took to be modal and fallacious, and on Tarski's appeals to the "common concept of consequence".

§1. Introduction. In the late 1980s, John Etchemendy, as a part of his well-known critique of the model-theoretic account of logical consequence, provided a historical description of Tarski's original proposal of the account in 1936.[1] If one accepted this description, one had to acknowledge that the grounds Tarski had given for the account were problematic or fallacious, and that, against what many readers of Tarski had thought, some of his key theoretical presuppositions were eccentric and defective. This reinforced Etchemendy's purely philosophical critique, for on his description some of the main later endorsements of the account or of aspects of it[2] turned out to rely either on the faulty Tarskian arguments or on a mistaken belief that some of Tarski's key theoretical presuppositions were the well-grounded ones that later would become standard.

There were three main groups of historical claims of Etchemendy's that had detrimental consequences for Tarski's 1936 proposal of the model-theoretic account. First, according to Etchemendy, when Tarski had said that one could give a proof that his account enjoyed a certain adequacy property he had called "condition (F)" (see below, section 4), he had meant in part that one could prove that if all the models of a set of sentences K are models of a sentence X (i.e., if X follows from K in Tarski's technical sense), then X follows by necessity from K; however, according to Etchemendy there is no correct proof of this, and Tarski's attempted proof commits a plain modal fallacy. Second, according to Etchemendy, when Tarski had appealed to the ω-incompleteness of certain theories as the main reason for proposing the model-theoretic account, his thought was that ω-arguments in a typical first-order language for elementary arithmetic (arguments with a set of premises of the

Received: October 28, 2008

[1] Tarski's account is in Tarski (1936a) (English version, Tarski, 2002) and Tarski (1936b) (English version, Tarski, 1983b; quotations will generally be from this English version; in a few places where this version is susceptible of potentially substantial corrections, these will be explicitly noted); see Stroińska and Hitchcock (2002) for an account of the differences between Tarski (1936a) and Tarski (1936b). Etchemendy's historical description of Tarski appears mainly in Etchemendy (1988, 1990, chapter 6).

[2] See the next footnote for some references.

© 2009 Association for Symbolic Logic
doi:10.1017/S1755020309090200

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

forms $F0$, $Fs0$, $Fss0$, etc. and conclusion $\forall x Fx$) were logically correct; however, in order to be able to hold that all the models of $F0$, $Fs0$, $Fss0$, etc. are models of $\forall x Fx$, he made the eccentric assumption that 0 and s must have the same (intended) extensions, and the numerical quantifier $\forall x$ must have the same (intended) domain, across different interpretations of a first-order language for elementary arithmetic. Third, according to Etchemendy this view of ω-arguments in a typical first-order language for elementary arithmetic led Tarski to require all interpretations of any given first-order language to share the same domain of quantification, the domain of the intended interpretation of the language's quantifier (the set of natural numbers in the case of elementary arithmetic); however, this had the bizarre effect that a sentence of such a language like $\exists x \exists y(x \neq y)$, stating the existence of two different objects of the kind talked about in the language (two natural numbers in the case of elementary arithmetic), was true in all interpretations and hence a logical consequence of all sets of sentences (a logical truth) on Tarski's account.[3]

It seems fair to say that, after the wave of responses to Etchemendy in the 1990s by Greg Ray, Gila Sher, myself, and others,[4] Etchemendy's first two groups of claims are now regarded as at least very doubtful by a majority of scholars, although issues of interpretation subsist concerning the relevant Tarskian passages. However, despite the fact that the third group of claims was primarily introduced by Etchemendy as a consequence of the now discredited second group, a substantive number of scholars still accept (versions of) his view that Tarski required all interpretations of any given language to share one same domain of quantification. This paper's first aim is to show (in section 2) that recent defenses of this view by Paolo Mancosu and Timothy Bays are based on misinterpretations of Tarski; in particular, I will rebut their criticisms of my earlier attack on Etchemendy's fixed-domain exegesis and I will offer a much more detailed report of the textual evidence on the issue than in my earlier paper "Tarski on logical consequence" (Gómez-Torrente, 1996a; "TLC" henceforth). I will also offer new considerations on the issues of interpretation that subsist about Tarski's views on the logical correctness of certain ω-arguments (section 3) and on the proof that Etchemendy took to be modal and fallacious (section 4), discussing in particular criticisms of my earlier claims by Bays and David Hitchcock. Finally, I will offer some critical comments on an exegesis of Tarski's claims about the "common concept of consequence" recently presented by Ignacio Jané; this will prove a convenient entry point into a sketchy picture of what I take to be the main Tarskian pretheoretic ideas about logical consequence (section 5).

A theme that informs my approach to many of the issues of interpretation, already present also in TLC, is what I take to be Tarski's "pluralism", that is his disposition to work

[3] It is worth noting that several authors who later endorsed Tarski's 1936 model-theoretic account, or aspects of it, accepted the Tarskian grounds attacked by Etchemendy, and saw Tarski as presupposing that different interpretations of a given first-order language could have different domains of quantification. Thus Church (1956, §55, n. 533 and surrounding text) takes Tarski's account to codify the notion of consequence appropriate for mathematical theories from the point of view of the axiomatic tradition, and this involves a variable-domain notion of interpretation that Church explicitly attributes both to Tarski (1936a) and Tarski (1936b). Prawitz (1985) accepts that Tarski's ω-incompleteness considerations show the inadequacy of derivability characterizations of logical consequence. Carnap (1942, p. 87) sees as one essential virtue of the account in Tarski (1936b) the fact that it imposes a condition on logical consequence that is stronger than condition (F).

[4] Sher (1991), Gómez-Torrente (1996a), and Ray (1996) questioned Etchemendy's third group of claims, Gómez-Torrente (1996a) and Sher (1996) questioned the second, and Gómez-Torrente (1996a, 1998), Ray (1996), and Sher (1996) questioned the first.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

within a variety of different frameworks, each with its own assumptions and conventions. On topics of what later would come to be called model theory, we find Tarski ready to work under logicist assumptions and conventions, but also within a framework more suited to the ideas of the axiomatic tradition, and specifically under suppositions peculiar to the Hilbert school. This Tarskian pluralism bears significantly on his defined concept of logical consequence, which he sought to make applicable to the different frameworks in which he worked. I expect to show that a failure to appreciate Tarski's desire to accommodate these different frameworks is to blame for some mistakes in Tarskian exegesis.

While many of my remarks will be strongly critical, the ultimate purpose of the paper is the constructive one of making new steps toward a satisfactory account of Tarski's views on logical consequence. I hope it will be obvious to the reader that the authors criticized here have contributed significantly to the shaping of my positive views, which incorporate a number of their ideas. We are all together in this difficult, perhaps unexpectedly difficult, interpretative task.

§2. Variable domains versus fixed domain.

2.1. Tarski's grammatical conventions for the use of variable domains. TLC began by pointing out that there was no reason to think that Tarski saw the ω-arguments in a typical first-order language for elementary arithmetic as logically correct. Instead, there was evidence that the ω-arguments he was thinking of as logically correct were arguments formulated in the simple theory of types, with the arithmetical constants defined in the logicist fashion and the numerical quantifier ranging over type-theoretical constructions that could represent the numbers (much more on this in section 3). Provided one accepts this, there is no need to postulate that Tarski held an eccentric view that the arithmetical constants of a typical first-order language for elementary arithmetic are logical constants and the first-order numerical quantifier is not subject to reinterpretation. The arguments with a set of premises of the forms $F0$, $Fs0$, $Fss0$, etc. and conclusion $\forall x Fx$ were simply not logically correct for Tarski, and there is thus no need to suppose that the numerical quantifier $\forall x$ had only one possible domain for him.

Nevertheless, Etchemendy had also observed that no apparent mention of possible changes in the domain of quantification was made by Tarski in his 1936 paper. For Tarski an interpretation of a formal language seemed merely to be a sequence of extensions assigned to the nonlogical constants of the language; Tarski did not seem to stipulate a special place for a set constituting the domain of quantification of the interpretation. Etchemendy had taken this as evidence for his claims about Tarski's understanding of ω-arguments, but if these claims were wrong, as now seemed clear, the interpreter had to give some explanation of the fact that Tarski did not explicitly mention possible changes in the domain of quantification across interpretations of a language, and preferably one that did not imply that Tarski thought that the sentence $\exists x \exists y (x \neq y)$ was a logical truth of first-order arithmetic.

In TLC I claimed that, for languages like a typical first-order language for elementary arithmetic, Tarski frequently (though not always) used at the time the conventions (C1) of naming the domain of quantification of the intended interpretation with a nonlogical predicate to which quantifiers must be relativized, and (C2) requiring the domains of quantification of alternative interpretations to be named by that same predicate; in particular, I claimed that Tarski employed (C1) and (C2) with the first-order languages used to formalize axiomatic mathematical theories in his manual (Tarski, 1937a; English version

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

with revisions and additions, Tarski, 1941).[5] TLC claimed that when he wrote his paper on logical consequence, Tarski probably had in mind these conventions for the typical first-order languages used in the formalization of axiomatic mathematical theories, that is languages with nonlogical mathematical constants and with quantifiers that are initially stipulated to range over a set of individuals of some specific mathematical nature (e.g., natural numbers, real numbers, etc.). If this is so, then Tarski viewed changes in the domain of quantification of a language of this kind as changes in the interpretation of a predicate, and the interpreter would have what seemed needed. Under (C1) and (C2), $\exists x \exists y(x \neq y)$ is just short for, say, $\exists x \exists y(N(x) \ \& \ N(y) \ \& \ x \neq y)$, and it is not true in all interpretations of its language. As I will stress later, this is compatible with claiming, as TLC did also claim, that Tarski did not require (C1) or (C2), not even in the context of the 1936 paper, for languages that were not strictly typical mathematical languages, for example languages without nonlogical constants and/or with "logical" quantifiers, quantifiers that are not initially stipulated to range over a set of individuals of a specific (e.g. mathematical) nature; and it is compatible with claiming, as TLC did also claim, that Tarski did not require (C1) or (C2) in other contexts of investigation, even for typical first-order mathematical languages.

Recently both Bays (2001) and, in greater detail, Mancosu (2006), have rejected even the weak claim that for many typical first-order mathematical languages Tarski employed the convention (C2), of representing domain changes as changes in the extension of a nonlogical predicate. They concede that Tarski used a nonlogical predicate to name, for example, the set of natural numbers in some languages, and even that in some cases Tarski may have taken the domain of quantification to coincide with the extension of a nonlogical predicate. But they claim that in no case did Tarski postulate that changes in the interpretation of that predicate covaried with changes in the domain of quantification. According to them, the domain of quantification is the same in all interpretations of a given language, and the possible extensions of the predicate in question are the subsets of that domain, which will thus properly include the extension of the predicate in many interpretations.[6]

---

[5] Tarski also used a related convention with a second-order language for real arithmetic in his manual (see below). My main focus will be on the first-order languages that were also the main focus of Etchemendy.

[6] Beyond this, there are striking differences between the fixed-domain views of Bays and Mancosu, and between both of their views and Etchemendy's. Etchemendy considered only first-order mathematical languages and claimed that for Tarski the intended domain of quantification of a language of this kind is the domain of quantification of all its possible interpretations. Unlike Etchemendy, Bays claims that Tarski contemplated domain variation in some contexts of investigation, but not when he gave his definition of logical consequence in the 1936 paper; however, he also differs from Etchemendy in claiming that, in the context of the 1936 paper, Tarski requires all interpretations, of all kinds of languages, to have as domain of quantification the class of "all" individuals (of the world). (Bays, 2001, pp. 1709–1710, wrongly attributes this view to Etchemendy, perhaps momentarily forgetting that the latter's main motivation for his fixed-domain exegesis is the misinterpretation of Tarski's claims about ω-arguments as claims about first-order arithmetic; see p. 1710, n. 11, where this is not even mentioned in a list of conjectured Etchemendyan grounds for the fixed-domain exegesis.) Mancosu, like Etchemendy and unlike Bays, claims that for Tarski the intended domain of quantification varies from language to language (and that the intended domain provides the range of quantification for all the interpretations of the language), but he makes the claim for every language considered by Tarski. (Mancosu, 2006, p. 232, wrongly attributes to Etchemendy the idea that the domain of the quantifiers is in all cases the domain of an underlying type theory.)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

As far as I can see, Bays does not offer any textual argument for the negative claim that Tarski did not use convention (C2) in his manual.[7] He just adduces the consideration that "in general, relativisation to a predicate doesn't *do anything* for a variable-domain conception of models: it amounts to little more than a (gratuitously complicated) way of adding the axiom '$\forall x D(x)$' to our theory" (Bays, 2001, p. 1712, n. 19; 'D' is the "domain predicate"). This is clearly unsatisfactory in this context, and not just because it is not a textual argument. The effect of the conventions I attributed to Tarski in his manual, and in particular of (C2), is stronger than the effect of merely "adding the axiom '$\forall x D(x)$' to our theory"; (C2) constrains the domain of quantification to vary along with the interpretation of the domain predicate, while the mere presence of an axiom of the form '$\forall x D(x)$' in a theory not subject to convention (C2) is indeed compatible with a fixed domain for all the interpretations of its language (see note 24 and surrounding text below).

Mancosu, on the other hand, is well aware of the effect of the conventions postulated in TLC, but does claim that it can be decisively established on textual grounds that the examples in Tarski's manual do not employ the mentioned convention (C2). He notes that the convention "would force the predicate S [naming the domain of quantification] to have a non empty extension. In fact $\exists x(x = x)$ is a theorem of the background logic and thus the [convention] would always force the theory T to have as an axiom $\exists x(S(x) \,\&\, x = x)$" (Mancosu, 2006, p. 220). However, he also claims that

> it is not hard to show that Tarski would not accept [the consequence in question]. On p. 145 of his 'Einführung in die mathematische Logik' [Tarski, 1937a], Tarski is discussing a theory $\mathfrak{A}$ formulated by means of the predicate Zl (for Zahl) and he remarks that the theory does not prove or disprove the statement $\exists x\mathrm{Zl}(x)$. Since, as I pointed out, under Gómez-Torrente's [convention] $\exists x(\mathrm{Zl}(x) \,\&\, x = x)$ (and thus $\exists x\mathrm{Zl}(x)$) would have to be an axiom (resp. a theorem) of the theory, it follows that the constraint is incorrect when judged against Tarski's practice (Mancosu, 2006, p. 220).

It will be important to begin by showing that Mancosu's claim is wrong and, furthermore, that examination of the textual evidence relevant to the claim actually establishes that Tarski was employing the convention in question.

The theory $\mathfrak{A}$ mentioned by Mancosu is a formalization of a weak fragment of the elementary arithmetic of the real numbers. Its language has the symbols Zl (a one-place predicate), $<, >$ (two two-place predicates), and $+$ (a two-argument function symbol) as mathematical primitives, besides the usual logical symbols of the first-order predicate

[7] In section 3, in the context of our discussion of ω-arguments, we will examine Bays's basic argument for the positive claim that in the 1936 paper Tarski assumed that all interpretations of every language have as domain the class of "all" individuals of the world. Another negative argument Bays gives is against the claim, that he attributes to some of Etchemendy's critics, that "fixed-domain conceptions of models (...) have disastrous consequences for mathematical practice" (2001, p. 1712) and hence it would be uncharitable to attribute a conception of that kind to Tarski. Bays points out that standard theorems that presuppose the variable domains conception standard nowadays can be given more complicated equivalent formulations that presuppose fixed-domain conventions. But no claim to the contrary was made in TLC, whose arguments were almost without exception not "charity" arguments, but arguments based on purely textual grounds. (I find perplexing Bays's earlier claim (at p. 1702, n. 1), that the arguments in TLC rest "largely on grounds of charity".)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

calculus (including identity) and an appropriate supply of individual variables.[8] Its only remarkable grammatical feature from the point of view of current conventions is that Tarski requires all quantifications to be relativized to the predicate Zl. $\mathfrak{A}$ has the following 11 axioms:

1. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset x = y \vee x < y \vee x > y)$;
2. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset (x < y \supset\sim y < x))$;
3. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset (x > y \supset\sim y > x))$;
4. $\forall x \forall y \forall z(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, \mathrm{Zl}(z) \supset (x < y \,\&\, y < z \supset x < z))$;
5. $\forall x \forall y \forall z(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, \mathrm{Zl}(z) \supset (x > y \,\&\, y > z \supset x > z))$ (cf. Tarski, 1937a, p. 99).
6. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset \exists z(\mathrm{Zl}(z) \,\&\, x + y = z))$;
7. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset x + y = y + x)$;
8. $\forall x \forall y \forall z(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, \mathrm{Zl}(z) \supset x + (y + z) = (x + y) + z)$;
9. $\forall x \forall y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \supset \exists z(\mathrm{Zl}(z) \,\&\, x = y + z))$;
10. $\forall x \forall y \forall z(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, \mathrm{Zl}(z) \supset (y < z \supset x + y < x + z))$;
11. $\forall x \forall y \forall z(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, \mathrm{Zl}(z) \supset (y > z \supset x + y > x + z))$ (cf. Tarski, 1937a, p. 110).

Imagine now that we suppress the relativizations to the predicate Zl, so that the axioms look more like they would under present conventions. It is not hard to see that the unrelativized axioms have models with one-element domains. In particular, the interpretation where the domain is {0} and the extensions of $<$, $>$, and $+$ are $\emptyset$, $\emptyset$, and $\{(0, 0, 0)\}$, respectively, is one such model. Therefore, given the standard first-order logic, the theory does not imply the existence of more than one number. If Tarski was using the convention that Zl named the domain of quantification in every interpretation, then neither do the relativized original axioms imply the existence of more than one number, but, as Mancosu notes, they imply existential sentences like $\exists x(\mathrm{Zl}(x) \,\&\, x = x)$ and others.

What does Tarski say in the passage referred to by Mancosu? Tarski has just said that while a strict (*streng*), or finitist, proof of consistency for stronger theories of the arithmetic of the real numbers is impossible, a consistency proof of this kind for theory $\mathfrak{A}$ is easy for the specialist. He then says that this is due to the fact that "on the basis of this system [$\mathfrak{A}$] it is not possible, for example, to answer positively or negatively the question whether there are *numbers* at all" (Tarski, 1937a, p. 145, my emphasis). This is the passage alluded to by Mancosu. Shortly afterward, Tarski adds that the finitist proof of consistency would become impossible merely if one added to $\mathfrak{A}$ a single sentence stating that "there are at least *two different numbers*" (Tarski, 1937a, p. 146, my emphasis), as this would imply, with the help of the other axioms, the existence of infinitely many numbers. It is thus fairly obvious that for Tarski the question that, being undecidable in $\mathfrak{A}$, makes a finitist proof of consistency for $\mathfrak{A}$ possible, is the question as to whether there are at least *two* different numbers; hence the use of the plural in the first quoted passage.[9] Note that Tarski could not have claimed that what makes a finitist proof of consistency for $\mathfrak{A}$ possible is the alleged

[8] Later Tarski extends $\mathfrak{A}$ to a very strong second-order theory of real arithmetic considered below, $\mathfrak{A}^{\times\times}$. This abides also by a convention related to (C2): second-order quantifications are relativized to the predicate '$\subseteq$ Zl' (see Tarski, 1937a, p. 154).

[9] The Polish original (Tarski, 1936c), at p. 148, also uses the plural.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

fact that $\exists x \mathrm{Zl}(x)$ is not decidable in $\mathfrak{A}$, for even if this sentence were added as an axiom and thus explicitly decided in the positive, the resulting system would have an easy finitist consistency proof.[10]

Mancosu has probably been misled here because the sentence "it is not possible to answer positively or negatively the question whether there are numbers at all", taken out of context, admits, despite of the plural, a purely existential reading ("it is not possible to answer positively or negatively the question whether there is any number at all").[11] Thus it is not out of order to exhibit further textual evidence that makes it even clearer that Tarski thought of $\exists x \mathrm{Zl}(x)$ as provable in $\mathfrak{A}$. On p. 131 of the manual, Tarski proposes the following exercise for the reader:

> 24. In section 48 we have given the definition of the symbol '0' by way of example. If one wishes to have the certainty that this definition does not lead to contradiction, one must have first the following theorem: *There is exactly one number z such that, for any number x, the formula:* $x + z = x$ *holds*. Prove this sentence on the basis of Axioms 6-9 [of $\mathfrak{A}$] alone. (Tarski, 1937a, p. 131)

It is thus clear that Tarski thought that the sentence $\exists z(\mathrm{Zl}(z) \,\&\, \forall x(\mathrm{Zl}(x) \supset x + z = x) \,\&\, \forall y(\mathrm{Zl}(y) \,\&\, \forall x(\mathrm{Zl}(x) \supset x + y = x) \supset z = y))$, which licenses the introduction of '0' as a defined symbol, was a theorem of $\mathfrak{A}$. From this sentence follow, under any view, Mancosu's sentences $\exists x(\mathrm{Zl}(x) \,\&\, x = x)$ and $\exists x \mathrm{Zl}(x)$.

If Tarski was not using the convention that Zl names the domain of quantification in every interpretation of the language of $\mathfrak{A}$, and was instead using the conventions proposed by Mancosu and Bays (that the domain of quantification is the same in all interpretations of the language of $\mathfrak{A}$, and the possible extensions of Zl are the subsets of that domain), Zl would have an empty extension in some interpretations of that language. However, note that it would also have an empty extension in some trivial models of $\mathfrak{A}$. (Mancosu himself points out that on his reading of Tarski $\exists x \mathrm{Zl}(x)$ is not a consequence of $\mathfrak{A}$.) Since clearly Tarski thought that $\exists x \mathrm{Zl}(x)$ was a theorem of $\mathfrak{A}$, it follows that he was using the convention that Zl names the domain of quantification in every interpretation of the language of $\mathfrak{A}$.

In fact, many other things Tarski says show that for him the interpretations of theories like $\mathfrak{A}$ may have different (and even disjoint) domains of quantification, which are always the extensions of the nonlogical "domain predicate". It will be good to devote some space to a review of these texts. A serious defect of TLC was that it did not review these texts

---

[10] Tarski says that although the finitist proof that $\mathfrak{A}$ is consistent is easy, it would be impossible to give it without fully formalizing the theory, which is outside the bounds of his manual. The proof would presumably consist in observing (1) that the axioms of $\mathfrak{A}$ are all satisfied in a finitary construction such as the one-membered model above; (2) that the formalized rules of inference preserve the property of being satisfied in this construction; and (3) that if some formula F is satisfied in the construction, then $\sim F$ is not so satisfied. Note that $\exists x \mathrm{Zl}(x)$ is satisfied in the finitary construction in question, assuming of course that Zl has {0} as its extension.

Both Mancosu and Bays claim that when Tarski uses relativized axiomatizations such as $\mathfrak{A}$, he embeds them in a suitable type theory, often with the axiom of infinity. It is thus worth noting that Tarski cannot have thought of the logic of $\mathfrak{A}$ as being provided by a strong type theory, for surely he then could not have claimed that $\mathfrak{A}$ has a finitist consistency proof.

[11] The translator of the English edition (O. Helmer) was similarly misled, as he translated: "it is (...) not possible to give (...) an answer to the very elementary question as to whether *any* numbers exist at all" (Tarski, 1941, p. 205, my emphasis; the 'any' is lacking both in the Polish original and in the German translation (by Tarski)).

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

in any detail, and only pointed to them; a review would probably have prevented some misunderstandings.

Let's begin by noting that Tarski (1937a) is completely explicit that arithmetic (by which in this context he means real number arithmetic, or analysis), as he understands it in his manual, is a theory whose variables *range exclusively over numbers*:

> Of the variables "x", "y", ... that are used in arithmetic one says that they *stand for designations of numbers*. With this one means approximately the following: a sentential function containing the symbols "x", "y", ... becomes a sentence when one replaces these symbols with constants that designate numbers (and not with expressions that designate operations on numbers, or relations between numbers, or any things outside the field of arithmetic such as geometrical figures, animals, plants, etc.) (Tarski, 1937a, p. 4)

(The slightly expanded translation Tarski, 1941, at p. 6, adds "or that numbers are VALUES of these variables" after "they *stand for designations of numbers*".) This shows beyond reasonable doubt that Bays's claim that in Tarskian relativized axiomatizations the domain of quantification is the class of all individuals of the world is false. It also goes against Mancosu's unqualified claim that "Tarski thinks that most theories (arithmetic, analysis, geometry) are such that the identity [between the intended domain of the theory and the range of the quantifiers] does not hold" (2006, p. 226). Tarski (1941) is also completely clear that quantified arithmetical sentences superficially not relativized to a nonlogical predicate in fact abbreviate relativized sentences:

> For the sake of brevity, the quantifiers are sometimes replaced by symbolic expressions. We can, for instance, agree to write in place of:
>
> *for any things (or numbers) x, y, ...*
>
> and
>
> *there exist things (numbers) x, y, ... such that*
>
> the following symbolic expressions:
>
> $$\underset{x,y,\ldots}{\mathbf{A}} \quad \text{and} \quad \underset{x,y,\ldots}{\mathbf{E}}$$
>
> respectively. (Tarski, 1941, p. 10)

When Tarski (1937a) defines the notion of model of a deductive theory, he does so by means of an illustration that involves a "miniature theory" about the congruence of segments. He is again completely explicit that here the variables range exclusively over segments, and that a nonlogical predicate names the domain of the intended interpretation:

> Imagine that we are interested in general theorems about the congruence of segments and that we intend to construct this fragment of geometry as a special mathematical theory. *We therefore prescribe that the variables "x", "y", "z"... designate segments*. As basic concepts we choose the signs "S" and "$\cong$"; the first of them is an abbreviation of the expression "the set of all segments", the second designates the relation of congruence. The formula:
>
> $$x \in S$$

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> is also read in the following way: "x belongs to the set of all segments", or simply: "x is a segment"; *the formula:*
>
> $$x \cong y$$
>
> *says that the segments x and y are congruent*. (Tarski, 1937a, pp. 84–85, my emphases)

Also, all the quantifications in the axioms of this miniature theory and in the theorems that Tarski proves by way of illustration are relativized to the predicate 'S'. For example, the axioms are these two:

> **Axiom I.** If $x \in S$, then $x \cong x$ (in other words: every segment is congruent to itself).
> **Axiom II.** If $x \in S, y \in S, z \in S, x \cong z,$ *and* $y \cong z$, then $x \cong y$" (in other words: two given segments congruent to a third are congruent to each other). (Tarski, 1937a, p. 85)

Tarski (1937a), at pp. 81–82, gives a general definition of the notions of interpretation and model of an axiom system, using the procedure and words very similar to those in Tarski (1936b), thus reinforcing the hypothesis that he assumes the same conventions in both places. First one turns the axioms into sentential functions, replacing the primitives of the theory with suitable variables. One can then consider the question whether a sequence of things satisfies these sentential functions. (Tarski, pedagogically, says that this is equivalent to the question whether the sentential functions become true when one replaces the variables with names of the appropriate things of the sequence.) When a sequence of things satisfies the sentential functions that the axioms give rise to, one says that the things in the sequence order form a model of the given axiom system. Tarski (1941), probably for pedagogical reasons, doesn't give general definitions and illustrates everything from the beginning using the "miniature theory" about segments:

> Let us replace the primitive terms in all axioms and theorems of our theory by suitable variables, for instance, the symbol '**S**' by the variable 'K' denoting classes, and the symbol '$\cong$' by the variable 'R' denoting relations (. . .). The statements of our theory will then be no longer sentences, but will become sentential functions which contain two free variables, 'K' and 'R', and which express, in general, the fact that the relation R has this or that property in the class K. (Tarski, 1941, p. 122)

> If a relation R is reflexive and has the property **P** in a class K [i.e., that for any elements x, y and z of the class K, if xRz and yRz, then xRy], we say that K and R together form a *model* or a *realization of the axiom system* of our theory, or, simply, that they satisfy the axioms. (Tarski, 1941, p. 123; see p. 122 for "property **P**").

That the interpretations of theories like the miniature theory may have different (and even disjoint) domains of quantification is already fairly clear from the examples of models of the miniature theory given in Tarski (1937a), pp. 85–86, and Tarski (1941), pp. 123–125. The first example of a model of the theory, of course, has as denotation of 'S' the set of all segments, and as denotation of '$\cong$' the congruence relation among segments. The second example has as denotation of 'S' the set of all things, or logical individuals, and as denotation of '$\cong$' the universal identity relation among all things. The third example has

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

as denotation of 'S' the set of real numbers, and as denotation of '$\cong$' the relation, called "equivalence", in which two real numbers x and y stand when their difference $x - y$ is an integer (cf. Tarski, 1937a, p. 86). Tarski also notes that a model of the miniature theory is "formed by an arbitrary class of individuals and by the universal relation V holding between any two individuals" (Tarski, 1941, p. 125).

Take any model $(\mathbf{A}, \mathbf{A} \times \mathbf{A})$ of this kind with $\mathbf{A}$ disjoint from the class of segments $\mathbf{S}$. Since $\mathbf{A}$ is disjoint from $\mathbf{S}$, it is not included in the domain of quantification of the intended interpretation of the theory. If we interpret 'S' and '$\cong$' by means of $(\mathbf{A}, \mathbf{A} \times \mathbf{A})$ and assume, as we must on the fixed-domain exegesis, that variables must still range over segments, which explicitly constitute the domain of quantification of the intended model for Tarski, we quickly get into trouble. For example, the sentence $\forall x(S(x) \supset\sim x \cong x)$ is vacuously true in $(\mathbf{A}, \mathbf{A} \times \mathbf{A})$ if 'x' ranges over segments (since 'S' denotes $\mathbf{A}$). But surely its negation $\sim \forall x(S(x) \supset\sim x \cong x)$ (or $\exists x(S(x) \,\&\, x \cong x)$)) is a theorem of the theory, as we know that $\exists x S(x)$ is a logical theorem for Tarski. So, absurdly, both $\forall x(S(x) \supset\sim x \cong x)$ and its negation should be true in $(\mathbf{A}, \mathbf{A} \times \mathbf{A})$.

In order to confirm the claim that the interpretations of the theories in the manual may have disjoint domains of quantification, we don't even have to use the otherwise clear fact that Tarski takes sentences like $\exists x S(x)$ as logical theorems. Tarski says that a certain geometrical interpretation $(\mathbf{P}, \dot{<}, \dot{>}, \dot{+})$, where $\mathbf{P}$ is the set of points of any given straight line and $\dot{<}$, $\dot{>}$, and $\dot{+}$ are two suitable two-place relations on $\mathbf{P}$ and a two-argument operation defined on $\mathbf{P}$, respectively, "forms a model of the mentioned axiom system [the theory of arithmetic $\mathfrak{A}$ above]" (1937a, p. 128). But he also says that "the points of this straight line are designated with the letters 'x', 'y', 'z', ..." (1937a, p. 127). Thus it is clear that the variables of the language of $\mathfrak{A}$ come to range, under the geometrical interpretation, over the points of the straight line that constitute the extension of the domain predicate Zl in the geometrical interpretation.

There is more evidence, and just as decisive. Tarski (1937a, pp. 157–158) expands $\mathfrak{A}$ to a stronger, second-order theory of the arithmetic of the real numbers $\mathfrak{A}^{\times\times}$, based on a somewhat different language with 'Zl', '$<$', '$+$', '0', '$\cdot$' and '1' as nonlogical primitives (in the language of $\mathfrak{A}^{\times\times}$, second-order quantifiers are relativized to the predicate '$\subseteq$ Zl', and first-order quantifiers are still relativized to 'Zl'). The axioms of $\mathfrak{A}^{\times\times}$, unlike those of $\mathfrak{A}$, imply directly the existence of two different numbers: 'Zl(0)', 'Zl(1)' and '$0 \neq 1$' are all axioms of $\mathfrak{A}^{\times\times}$; hence the sentence $\exists x \exists y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, x \neq y)$ is an immediate theorem of the theory, and so is any weaker existential sentence like $\exists x \mathrm{Zl}(x)$ (even without appealing to the fact that this was already a first-order logical theorem). Tarski gives as an exercise for the reader to extend the geometrical interpretation mentioned in the preceding paragraph to a geometrical interpretation which is a model of $\mathfrak{A}^{\times\times}$ (1937a, p. 163, exercise 9). Let $(\mathbf{P}^{\times\times}, \dot{<}^{\times\times}, \dot{+}^{\times\times}, \dot{0}^{\times\times}, \dot{\times}^{\times\times}, \dot{1}^{\times\times})$ be one such interpretation. The same trouble as above arises for the fixed-domain exegesis, now without using the assumption that $\exists x \mathrm{Zl}(x)$ is already a logical theorem. For example, the sentence $\forall x(\mathrm{Zl}(x) \supset x \cdot 1 \neq x)$ is vacuously true in $(\mathbf{P}^{\times\times}, \dot{<}^{\times\times}, \dot{+}^{\times\times}, \dot{0}^{\times\times}, \dot{\times}^{\times\times}, \dot{1}^{\times\times})$ if 'x' still ranges over the reals (since 'Zl' denotes now a set of points). But surely its negation $\sim \forall x(\mathrm{Zl}(x) \supset x \cdot 1 \neq x)$, or $\exists x(\mathrm{Zl}(x) \,\&\, x \cdot 1 = x)$, is a theorem of $\mathfrak{A}^{\times\times}$, as we know that $\exists x \mathrm{Zl}(x)$ is a theorem of $\mathfrak{A}^{\times\times}$ and $\forall x(\mathrm{Zl}(x) \supset x \cdot 1 = x)$ is an axiom of $\mathfrak{A}^{\times\times}$ (axiom 19 on p. 158). Hence on the fixed-domain exegesis we would easily get the absurd result that both $\forall x(\mathrm{Zl}(x) \supset x \cdot 1 \neq x)$ and its negation are true in $(\mathbf{P}^{\times\times}, \dot{<}^{\times\times}, \dot{+}^{\times\times}, \dot{0}^{\times\times}, \dot{\times}^{\times\times}, \dot{1}^{\times\times})$ from explicit claims of Tarski.

The reviewed textual evidence leaves few if any doubts that, for a typical first-order mathematical language, Tarski at least sometimes used a pair of conventions, (C1) and

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

(C2), that allowed interpretations with different domains of quantification for that language. There is also, I think, little doubt that Tarski presupposed these conventions in his 1936 paper, as preconditions for the applicability of his definition of logical consequence (as given in the paper) to typical first-order (and in general finite order) languages used in the formalization of axiomatic mathematical theories. Otherwise it would be difficult to understand Tarski's (1936b) claim that "in just this sense one usually speaks of models of an axiom system of a deductive theory" (p. 417), just after introducing the notion of a model using virtually the same words as in Tarski (1937a). It would be hard to understand how Tarski could introduce a nonstandard notion of model in a logic manual.

Both Etchemendy (1988, p. 73) and Mancosu (2006, p. 236) express a mild worry over the fact that their views imply that Tarski, at some point in the late 1940s or early 1950s, moved from a fixed-domain conception to a variable-domains conception, and yet there is no known record of the reasons he may have had for this.[12] Of course no such worry arises for my exegesis, but there are other related considerations that should make Etchemendy's and Mancosu's worry less mild. Tarski (1953) refers the reader again to Tarski (1935a, 1936b) "for formal definitions and a detailed discussion of semantical notions (satisfaction, truth, logical consequence, logical truth)" (Tarski, 1953, p. 8, n.7), thus implying that he has not withdrawn the definitions of the 1930s. Also, Tarski (1965), the "third edition" of the English version of Tarski's manual, was published well into the 1960s, and yet it's almost indistinguishable from the "second edition" (Tarski, 1946) and the first edition (Tarski, 1941). It is unreasonable to believe (as Etchemendy and Mancosu must) that Tarski, after having embraced the variable-domains conception, still republished a basic manual that was based on the fixed-domain conception. Tarski probably thought, even into the 1960s, that the use of (C2), in allowing him to define an interpretation simply as a sequence of extensions for the nonlogical constants, provided a more straightforward and pedagogical approach than the later standard one (of "unnamed" domains).

2.2. Tarski's use of fixed domains, and his alternative conventions for variable domains. The claim that Tarski sometimes used variable-domain conventions (C1) and (C2) is relatively weak. It is compatible with Tarski not having required (C1) or (C2), not even in the context of the 1936 paper, for languages that were not typical mathematical languages, for example languages without nonlogical constants and/or with "logical" quantifiers; and it is compatible with Tarski not having required (C1) or (C2) in other contexts of investigation, even for typical first-order mathematical languages. The variable-domains view, as defended in TLC, is only the weak view that Tarski had in mind a way of *allowing for* domain variation that was compatible with the formulation of his theory of logical consequence in Tarski (1936b), and that what he had in mind in this case was the use of (C1) and (C2). Mancosu, however, has taken my claims to be much stronger, and in particular has attributed to TLC the manifestly mistaken idea that Tarski did not merely presuppose (C1) and (C2) in suitable contexts, but also in *every* work he wrote in this period. According to Mancosu, I "claim that Tarski *always* pre-supposes that there will be a unary predicate characterizing the domain of the theory" in the works of this period (Mancosu, 2006, p. 225, my emphasis), and "the original strategy by Gómez-Torrente

[12] The dates are because Tarski (1941) (and its almost indistinguishable "second edition" Tarski, 1946) is supposed to hold still a fixed-domain conception (it is virtually the same text as Tarski, 1937a, which is strictly contemporaneous with Tarski, 1936b), and Tarski (1953) gives the model-theoretic definition of logical consequence for first-order languages using present conventions, including an explicit mention of varying an "unnamed" domain of quantification.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

(. . .) forces *in every theory* the identity between the intended domain of the theory and the range of the quantifiers" (Mancosu, 2006, p. 226, my emphasis). Mancosu documents how Tarski considered theories in which there is no nonlogical predicate characterizing the domain of quantification of the theory, or in which the intended domain of mathematical objects of the theory is properly included in the range of the quantifiers. But his effort is misdirected. TLC did not make any claim to the contrary, and in fact, as we will see, explicitly alluded to the existence of such theories considered by Tarski. However, TLC obviously failed to make clear the exact scope of my claims, which I will now attempt to make precise. This will also allow me to correct a secondary suggestion made in TLC, on Tarski's views about the relation of logical consequence in languages where the vocabulary of the simple theory of types is supplemented with mathematical nonlogical constants; here an unpublished lecture, Tarski (1940), unearthed by Mancosu, will be the main piece of evidence. But this correction of a tangential claim of TLC will not affect the variable-domains thesis, which as I said I take to be firmly established through the considerations of subsection 2.1.

When Etchemendy provided his 1988 description of Tarski's 1936 theory of logical consequence, he only attempted to cast doubts on the suitability of Tarski's theory to the task of characterizing the notion of logical consequence for typical first-order mathematical languages. Higher-order theories without mathematical primitives, or even higher-order theories with mathematical primitives, were simply not considered by him. And Etchemendy claimed that Tarski's theory mischaracterized the notion of logical consequence because it declared a sentence like $\exists x \exists y(x \neq y)$ a logical truth of first-order arithmetic.[13] If Tarski's theory had this implication, then it certainly was eccentric in a clear way. As TLC pointed out, "it is intuitively not the business of the underlying logic to decide how many individuals of a certain nature (natural numbers, real numbers, points) there are" (p. 142); "sentences meaning 'there are two natural numbers', 'there are two real numbers', and so on, should not be logical truths of the underlying logic for axiomatic first-order theories of the natural numbers, the real numbers and so on" (p. 143). The main purpose of TLC's claims concerning the issue of domain variation was, accordingly, to show that Tarski, in his 1936 paper, presupposed a convention that allowed interpretations with different domains of quantification for a typical first-order language used to formalize an axiomatic first-order theory (e.g. of the natural numbers, the real numbers, and so on).

But TLC was in no way committed to the claim that Tarski *always* employed this convention in the works of this period, or even to the weaker claim that the convention was taken by Tarski as a prerequisite for the application of his definition of logical consequence to languages that were not typical first- (and finite-) order mathematical languages. In fact, TLC rejected these two claims. I emphasized, for example, that Tarski (1936b) clearly wanted to apply his definition directly to the language of a certain formalization of simple type theory, the "General Theory of Classes" (GTC) of Tarski (1935a) (described in TLC, pp. 131–132, mentioned above in connection with ω-arguments, and examined at length below in section 3; see a compressed description in the accompanying footnote[14]). This is a theory without a single nonlogical constant that therefore obviously does not abide by convention (C1), let alone (C2). TLC devoted considerable space to showing that in

[13] And in general of mathematical languages whose intended domain of quantification was a domain of objects of a specific nature containing two different objects of that nature.

[14] Here is a quick rough description of GTC that only differs from Tarski's in that it uses notation now more common:

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

GTC Tarski, in consonance with much of the logicist tradition in which this theory was especially used, was ready to accept sentences like $\exists x \exists y (x \neq y)$,[15] and even an axiom of infinity expressed only with logical constants, as (type-theoretic) logical truths. The presumable reason was that the question whether there is a certain number of individuals of an *unspecific*, or "logical" nature, might be taken as a question that the underlying logic should decide, and that it should decide in the positive if the domain contains at least that number of individuals. (Of course it's also possible that Tarski's reason was a mere desire to follow established practice with languages of this kind.[16]) In Tarski's language for GTC there are no nonlogical constants, so if one applies to it his 1936 definition of logical consequence without any additional conventions, there is no variation of the domain of individuals across interpretations (there is in fact just one interpretation of the language[17]), and so, if the domain of "logical" individuals is infinite, those sentences are logically true. Thus Tarski's inclination to accept those sentences as logical truths in simple type theory is respected by his definition of logical truth (or consequence) (see TLC, p. 141).

---

a) primitive signs of GTC: $\forall$, $\vee$, $\sim$, (,), X, $'$, ' (the last two are a subindex accent and a superindex accent to generate variables by suffixing to 'X' $n(\geq 1)$ subindex accents and m superindex accents; the notation X_n^m is for the X followed by n subindex accents and m superindex accents).

b) grammar of GTC: Atomic formulas are of the form $X_n^{m+1} X_p^m$. Complex formulas are obtained by negation, disjunction, and universal quantification with respect to the variables of all orders.

c) interpretation of GTC: A variable X_n^1 takes as values individuals, a variable X_n^2 takes as values classes of individuals, and so forth. $X_n^{m+1} X_p^m$ "means" that the object assigned to X_n^{m+1} "contains" the object assigned to X_p^m. The other signs mean what you would expect.

d) deductive system of GTC: Some usual set of axioms and rules for the connectives and quantifiers, axioms of comprehension and extensionality for all orders, and an axiom of infinity.

[15] In the formulation of the preceding footnote, this would look like $\exists X_1^1 \exists X_2^1 (X_1^1 \neq X_2^1)$, with the first-order identity predicate defined with the help of second-order quantifiers.

[16] See the references to Ramsey (1925), Carnap (1934), and Hempel (1945) in TLC, p. 141, as examples of philosophers of a logicist persuasion who accepted the axiom of infinity as a logical truth in simple type theory. Even Hilbert and Ackermann (1928), p. 88, were ready to follow this practice, as noted in TLC, p. 139. See also the reference by Tarski to the distinction between the "general logical notion of individual" and specific notions of an individual (Tarski, 1941, p. 73), a distinction to be recalled below.

[17] It may be worth noting that Tarski must clearly be using some tacit convention that the language of GTC has at least one interpretation. It might seem that under Tarski's (1936a, 1936b) notion of interpretation (made somewhat more explicit in Tarski, 1940), the language of GTC has no interpretations, for an interpretation seems to be here a sequence of objects assigned to a nonempty set of nonlogical constants, and the language of GTC has no nonlogical constants. But if the language of GTC had no interpretations, then Tarski's definitions of logical truth and consequence for this language would be incorrect in a completely trivial way. For example, every sentence would be a logical truth, as it would be vacuously true that every interpretation of the language of GTC is a model of every sentence. For this reason, the most reasonable hypothesis, regardless of whether we sympathize with the variable-domains view or with the fixed-domain view, is to suppose that, in the case of a language without nonlogical constants, Tarski was working under the convention that at least the intended domain of quantification is suitably represented by some interpretation (perhaps the empty sequence), or constitutes by itself an interpretation of the language. All this is also confirmed by Tarski's claim (see subsection 2.3 below) that if all constants of a language are logical, his defined notion of logical consequence coincides extensionally with that of material consequence, and thus that the notion of logical truth coincides with that of truth.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

There were many other facts which TLC noted explicitly and which go against the strong claim that Tarski always employed the mentioned conventions in the works of this period. Thus, although TLC insisted on the canonical grammatical conventions that Tarski presupposed in (1936b) and (1937a) for handling domain variation, it also noted that for Tarski other conventions for this purpose were just as good, including some virtually indistinguishable from the now standard ones. Besides, clearly Tarski was aware that variable-domain notions of validity could be defined using his semantic apparatus, in such a way that they were directly applicable to languages that employed these other conventions. Let me now return to these claims with some more detail than in TLC.

In TLC I said that

> it can be amply documented that *many* of the theories containing mathematical primitives that Tarski formalizes (with the help of a logical basis without cardinality assumptions) explicitly contain extra-logical predicates which are true of all the objects in the domain of the intended interpretation (or interpretations, in algebraic theories) of the theory. (TLC, p. 144, my emphasis)

Note that I only said "many", not "all", and in fact a bit later on p. 144 and on note 13, p. 149, I mentioned explicitly a theory considered by Tarski that did not abide by the conventions in question, the first-order theory of the ordinal type ω with a sign $<$ for the natural relation of order among ordinals as the only nonlogical constant. For this theory, as noted in TLC, Tarski had proved the existence of denumerable models nonisomorphic to the intended model (see Editorial Office of *Fundamenta Mathematicae*, 1934). Of course, the fact that this theory and other first-order theories considered by Tarski (see especially Tarski, 1936d, §5 and appendix) do not abide by the conventions implicitly postulated in (1936b) and explicitly used in (1937a) implies that when studying these theories Tarski used a different convention, probably simply the standard one nowadays, according to which an interpretation consists of a possibly unnamed domain plus extensions for the nonlogical constants drawn from the domain. In fact, in the mentioned note (Editorial Office of *Fundamenta Mathematicae*, 1934), this exegesis is virtually inevitable, for it states a version of the upward Löwenheim–Skolem theorem as follows:

> Mr. Tarski has proved, as a supplement to the known theorem of Löwenheim–Skolem, that for every consistent set of first-order sentences that doesn't possess a finite interpretation, there exists a (not only countably infinite but) uncountably infinite interpretation. (Editorial Office of *Fundamenta Mathematicae*, 1934, p. 161)

If the authors (as we may suppose) accurately report Tarski,[18] this clearly presupposes that the word "interpretation" is being used by him in such a way that the domain may vary from interpretation to interpretation, even though in the languages mentioned in the note there is no "domain predicate".[19]

[18] In fact, it is natural to conjecture that Tarski, who had close ties with several of the editors of *Fundamenta Mathematicae* and probably did editorial work for the journal, was himself the author of the note or at least okayed it.

[19] This usage is at odds with Etchemendy's and Mancosu's view that Tarskian interpretations of a given language always have the same domain. The text also casts considerable doubts on Mancosu's speculation that Tarski must have viewed Löwenheim–Skolem theorems as complicated metatheorems about interpreted languages of type theory, such as (in the case above)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

TLC (pp. 139–140) also pointed out that Tarski considered favorably the possibility of defining logical validity for sentences of GTC more in line with the Hilbert school, as truth in all (nonempty) domains, thus contemplating a notion of logical consequence for sentences of simple type theory as preservation of truth in interpretations with variable domains of quantification (in this case, essentially, each domain constitutes an interpretation). He compares this notion with the logicist custom of taking the question of how many individuals of an unspecific nature exist to be a logical one:

> Of course, whoever wishes to transform the system of provable sentences of each investigated science into a complete one, must add to the system sentences that decide beforehand in one way or the other the question how many individuals there are. However, several circumstances make another point of view appear to be better founded, namely to leave the decision on such problems to the special deductive sciences, and in logic and its fragments to try to achieve only that the concept of provable sentence coincides as to its extension with the concept of a sentence correct in every domain of individuals. (Tarski, 1935a, p. 362)

When Tarski defines the notion of truth in every nonempty domain in Tarski (1935a), p. 320,[20] he does so directly, without using the convention that a nonlogical predicate names the domain of the intended interpretation. Thus we know for certain that in some contexts Tarski explicitly contemplated domain variation without convention (C2), although naturally his theory of logical consequence as formulated in Tarski (1936b) could only accommodate domain variation through some additional convention in the style of the convention (C2) used in Tarski (1937a). Of course, as TLC noted, Tarski generally stuck to the standard idea at the time, that a sentence of GTC is logically true just in case it is true in the domain of all individuals of the intended interpretation. But he was perfectly aware of and sympathized with the notion of truth in every nonempty domain as a possible notion of logical truth for sentences of GTC, and even defined it using his semantic apparatus.

Tarski (1935a), p. 318, n. 48, also notes that the Hilbert school employed a variable-domain notion of validity ("general validity") applied to formulas with free second-order variables (in the places where we now put nonlogical predicates) and with all first-order variables bound. These formulas are functionally equivalent to the sentences of first-order languages under present conventions. A formula of this kind is said to be valid in a domain when all the sequences of assignments (drawn from that domain) to the second-order variables satisfy the formula, and said to be generally valid when it is valid in every domain. Thus the notion of general validity for formulas of this kind is functionally equivalent to the now standard notion of validity for first-order sentences. Note that the basic definition for all these notions is that of the satisfaction of a formula by a sequence of

---

"For every set K of first-order sentences, if we have an interpreted language of type theory L' with an infinite domain and such that there is a model O' of K living in L', then there is an interpreted language L'' of type theory with an uncountably infinite domain and such that there is a model O'' of K living in L''" (cf. Mancosu, 2006, p. 235). Note that the text reporting Tarski's result attributes cardinality directly to the interpretations (and models) of the set of first-order sentences.

20 For simplicity of illustration he defines this notion for the "language of the calculus of classes", a simple second-order fragment of GTC where there is a (defined) logical constant for the relation of inclusion between subclasses of the individual domain. But for Tarski in 1935 it is of course perfectly obvious how to define the notion for larger fragments of GTC and even for GTC itself (see e.g. Gómez-Torrente, 2004).

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

objects drawn from a domain, with the objects providing extensions for the second-order variables; and Tarski is obviously implying that these notions can be defined using his methods, as he elsewhere says that the existing theorems about these Hilbertian notions and other semantic notions only "obtain a clear and distinct content" (Tarski, 1935a, p. 363) if one applies his method of definition. So in this context Tarski is again contemplating a variable-domain notion of logical truth, in this case with conventions essentially identical to present conventions; and he indicates that the notion can be defined using his semantic apparatus.

I take all this to show that not only was Tarski thoroughly aware of the variable-domain notion of validity for the language of simple type theory and for first-order mathematical languages, and that he used it substantially in his work (which we knew in any case from the texts considered in subsection 2.1), but also that he was aware of conventions for handling it which (unlike those he chose to presuppose in 1936b and 1937a) were essentially identical to present ones. Further, he was perfectly aware of how to define the variable-domain notion of validity in such a way that its applicability presupposed these conventions. TLC made no claim to the contrary, and even pointed briefly but explicitly to many of the facts we just revisited.

Another class of languages studied by Tarski that was briefly considered in TLC consists of those hybrid cases in which he "uses a simple type theory as what he calls a 'logical basis' for the formalization of several different mathematical disciplines" (TLC, p. 132); I said that

> in these formalizations (1) the domain of objects of the mathematical theory is either identified with the domain of individuals or taken to be defined by a primitive extra-logical predicate (*which does not apply to other members of the individual domain*); and (2) besides purely logical constants, there are some mathematical primitives denoting objects in the intended domain of mathematical objects or relations on the intended domain of mathematical objects. (TLC, pp. 141–142, my emphasis)

Thus I emphasized that Tarski used in some cases the formalism of the simple theory of types supplemented with nonlogical constants, to speak about a mathematical domain that is *properly* included in the domain of quantification of the type theory; note the phrase in italics. (See a description of a theory of this kind, from Tarski, 1931, in the accompanying footnote.[21]) In these cases, there was an extra-logical primitive (a nondefined

[21] Here is a quick rough description of the theory of the real numbers (RN) considered in Tarski (1931):

a) primitive signs of RN: $\upsilon, \leq, \sigma, \forall, \exists, \vee, \&, \sim, (,), \in, X, ', '$ (the last two are as in GTC).

b) grammar of RN: Atomic formulas are of the forms $X^m_n \in X^{m+1}_p$, $\upsilon(X^1_n)$, $X^1_n \leq X^1_p$, and $\sigma(X^1_n, X^1_p, X^1_q)$. Complex formulas are obtained by negation, disjunction, conjunction, and existential and universal quantification with respect to the variables of all orders.

c) interpretation of RN: A variable X^1_n ranges over a set containing individuals of an unspecific nature, among which are all the real numbers; a variable X^2_n takes as values classes of individuals; and so forth. $X^m_n \in X^{m+1}_p$ "means" that the object assigned to X^{m+1}_p "contains" the object assigned to X^m_n. $\upsilon(X^1_n)$ means that the object assigned to X^1_n is the number one; $X^1_n \leq X^1_p$ "means" that the object assigned to X^1_n is less than or equal to the object assigned to

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

symbol), whether logically atomic or not, that distinguished the specific subdomain,[22] but no nonlogical predicate stipulated to name the domain of quantification. So this is further proof that I did not see Tarski as having required "in every theory the identity between the intended domain of the theory and the range of the quantifiers".[23]

However, some of TLC's brief remarks about Tarski's view of the relation of logical consequence in languages with mathematical primitives based on type theory I now recognize as probably wrong, in view of Tarski's unpublished lecture "On the completeness and categoricity of deductive systems" (Tarski, 1940), unearthed by Mancosu. TLC says that "in these cases Tarski does not take any assumptions about the cardinality of the domain of individuals to form part of the underlying logic of the theory alone" (p. 142), relying on the fact that the theories considered in Tarski (1931) and Tarski (1935b) have a logical deductive basis without cardinality assumptions, consisting essentially of GTC minus the axiom of infinity (see again the description of the theory from Tarski, 1931, in note 21, and cf. Tarski, 1931, p. 213, and Tarski, 1935b, p. 81; see also Tarski and Lindenbaum, 1935, p. 15). This led me to think (though this was largely left implicit in TLC) that for languages of this kind Tarski adopted as his notion of logical validity the notion of truth in all nonempty domains (that as we saw he had thought to be "better founded" than the notion of simple truth in Tarski, 1935a). Or strictly speaking, the idea that a sentence S is logically valid when the "sentential function" S' that results from it, when one replaces the nonlogical constants in S with suitable new variables, is generally valid in the Hilbertian sense (i.e., true in all assignments to the new variables that can be drawn from any nonempty domain). This was (so I thought) one more of those cases in which Tarski brought into play domain variation without a convention like (C2).

But as Mancosu has noted (2006, p. 230), Tarski (1940) says that in languages of this kind, every sentence containing only logical constants and variables is such that either it or its negation is a logical consequence, in his 1936 defined sense, of the set of nonlogical axioms of the theory. Since presumably Tarski includes among the sets of nonlogical axioms he has in mind some weak sets without existential assumptions, and in general sets of nonlogical axioms that don't determine the exact number of individuals, this suggests that he thought that the intended domain of quantification in a theory of this kind settles all questions of cardinality, for example it settles whether sentences like $\exists x \exists y(x \neq y)$, or an axiom of infinity expressed only with logical constants and variables, are consequences of the theory or not. Thus it is suggested that the intended domain of quantification of a theory of this kind is the same in all interpretations of its language. Indeed, if the domain of all interpretations is the same, and it is, for example, infinite (e.g., because the subdomain of

---

X^1_p; $\sigma(X^1_n, X^1_p, X^1_q)$ "means" that the object assigned to X^1_q is the sum of the objects assigned to X^1_n and X^1_p. The other signs mean what you would expect.

d) deductive system of RN: Some usual set of axioms and rules for the connectives and quantifiers, axioms of comprehension and extensionality for all orders, and a usual set of axioms for the notions υ, $\leq$, and σ. Note that the logical axiom of infinity is not mentioned, presumably because it will be derivable from the axioms for the notions υ, $\leq$, and σ.

[22] For example, in the case of RN a predicate defining the set of real numbers is $\exists X^1_2(X^1_2 \leq X^1_1)$.

[23] In cases where the individuals of the specific nature to be studied did exhaust the domain of individuals in the intended interpretation, I did not say that Tarski required "that there will be a unary predicate characterizing the domain of the theory". But this will generally be so, as the mathematical domain will always be the extension of a predicate in the style of $\exists X^1_2(X^1_2 \leq X^1_1)$ for RN.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

mathematical objects is infinite), then the axiom of infinity will come out as a consequence of the set of nonlogical axioms (and in fact of every set of sentences, being thus a logical truth) on Tarski's definition (just as in the case of GTC reviewed above).

Thus Tarski is probably assuming a notion of logical consequence for these languages of simple type theory with nonlogical constants which is very similar to, in fact essentially the same as, the notion of logical consequence that he usually assumed for simple type theory without nonlogical constants. The implication is that in all these languages Tarski is happy to take the notion of an individual to be "logical", and thus that he is happy, following a certain established practice, to let questions of cardinality about the domain of individuals to be decided by logic. (Of course, questions about the cardinality of the intended subdomain of mathematical objects are still not decided by the underlying logic alone, as it should be.) In fact, in his manual Tarski implies that the "general logical notion of individual" can be employed in a theory with mathematical primitives, as he distinguishes between "specifying exactly what is considered an individual thing within the framework of [a particular mathematical theory]" and "using the general logical notion of individual within a particular mathematical theory" (Tarski, 1941, p. 73; he significantly adds that the former option "is sometimes more convenient" (as in Tarski, 1937a, or 1941)). Upon reflection, the assumption that the notion of an individual in these theories is "logical" seems at the very least a reasonable one to make, as their quantifiers, just as the quantifiers of GTC, are not initially stipulated to range exclusively over a set of individuals of a specific mathematical nature. These cases are not substantively different in this respect from the case of the simple theory of types without nonlogical constants discussed at length in TLC, and Tarski so took them to be.

(The same holds even for languages of this kind in whose associated theory there is an axiom of the form '$\forall x\, D(x)$' asserting that all the logical individuals are individuals of a certain nature. Tarski had recourse to axioms of this kind in Tarski (1935b) and in Tarski and Lindenbaum (1935), where he studies mathematical theories embedded in a type theory; but even in these cases that assertion is not an initial metatheoretical stipulation (as e.g. in the examples of Tarski, 1937a), but a mathematical assertion in the theory, designed to serve a certain practical purpose (see note 24). Note that there is no such axiom in the theories studied in the manual (even though the domain is stipulated to coincide with the extension of the "domain predicate"), because the specific intended domain of individuals is initially stipulated in the metatheory.[24])

[24] For the interested reader, Tarski's use for axioms of the form '$\forall x\, D(x)$' in mathematical theories embedded in type theory can be roughly described as follows. In Tarski (1935b) and Tarski and Lindenbaum (1935), such embedding plays a metamathematical role: many properties of the theories, such as their satisfiability in some model, their categoricity, and so forth, are enunciated with the help of quantifications over the higher types. (Tarski is just beginning to consider the use of set theory as a universal metatheory.) However, this entails some complications. Tarski wants to prove a number of theorems that involve the basic idea behind the concept of categoricity: under the notion usual at the time a set of sentences S is categorical when any two interpretations of its nonlogical constants that are models of S are isomorphic. Well-known second-order axiom systems R for analysis and N for arithmetic are categorical. Further, they are strictly categorical, in the sense that between any two models of $R(N)$ there is only one isomorphism. What Tarski wants to prove is that no strictly categorical axiom system A can be properly included in a categorical axiom system which contains nonlogical constants interpreted over the intended domain of A and which cannot be explicitly defined (with the help of quantifiers of any order) in terms of the nonlogical constants appearing in A. It is cumbersome to state and prove the theorem in type theory about constants with interpretations drawn from the intended domain of A. So

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

TLC's probable misdescription of Tarski's assumptions about the notion of logical consequence in a special kind of mathematical theories was the simple error of putting these theories in the wrong pigeonhole of the classification, not any substantive error in the structure of the classification itself. To recap: the variable-domains view (as expounded in TLC) is just the weak claim that Tarski contemplated a way of making domain variation compatible with the (1936b) formulation of his definition of logical consequence, not the false view that he required domain variation regardless of the context of investigation in play. The fact that Tarski did not employ domain variation to model the relation of logical consequence in some particular class of theories does not affect the view, as long as he employed domain variation to model that relation in some other class of theories (or in other contexts of investigation), in ways unequivocally compatible with the (1936b) formulation of his definition. Mancosu's criticism of TLC, based on the fact that some theories considered by Tarski in the relevant period did not abide by the conventions postulated in TLC as a precondition for the application of Tarski's definition to first-order mathematical theories, is based on a misunderstanding of the scope of the variable-domains view. The defender of the view does not need to defend, and TLC did not defend, the false claim that all the languages, or even all the first-order mathematical languages, considered by Tarski abided by those conventions. Defenders of the fixed-domain view like Mancosu and Etchemendy,[25] by contrast, have claimed that Tarski *required* domain invariance for every language, not just that he made use of a fixed domain in some contexts of investigation. Given the flexibility exhibited by Tarski's practice during this period of multiple alternative frameworks, assumptions, and conventions, it is not surprising that there are so many reasons, reviewed in this subsection and in subsection 2.1 above, why this strong view is flawed.

2.3. Logical consequence and material consequence. There is one final criticism of TLC's variable-domains view made by Mancosu and Bays, that cannot be left unaddressed. Mancosu summarizes it as follows (see also Bays, 2001, pp. 1708–1709):

> the great virtue of Etchemendy's position [on domain variation] is the natural account it yields of Tarski's claim, let us call it (LM), in the 1936

Tarski focuses on axiom systems A which, by virtue of an added axiom '$\forall x\, D(x)$' are true in an interpretation only if the intended domain of A coincides with the domain of the type theory. He also defines a corresponding notion of categoricity, absolute categoricity: an axiom system A is absolutely categorical when between any two interpretations of its nonlogical constants that are models of A there is an isomorphism that is also a permutation of the domain of individuals of the type theory. (Note that R is not absolutely categorical when it is embedded in type theory, if there is some individual a in the type-theoretic domain which is not a real number (as was compatible with the description of RN in note 21); consider the set r^* formed by the real numbers and a; there is clearly a model of R with r^* as domain, which is thus isomorphic with the intended model; but no isomorphism between these two models can also be a permutation of the universe of individuals, for a will be left out.) Then Tarski defines an axiom system to be strictly absolutely categorical when between any two models there is only one isomorphism that is also a permutation of the universe of individuals. What he ends up proving is that no strictly absolutely categorical axiom system A can be properly included in another absolutely categorical axiom system containing nonlogical constants which cannot be explicitly defined in terms of the nonlogical constants appearing in A. (Note that this is false when in place of 'absolutely categorical' one puts 'categorical'. We can add a nonlogical individual constant α to the language of R above, naming the individual a in the universe above, and add to R a sentence S saying that α is the one and only individual that is not a real number. $R + S$ is categorical but α cannot be defined in terms of the nonlogical constants in R.) This conveys in a simpler way the idea of the other more general theorem.

[25] But not Bays, as pointed out in a note 6.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> paper (p. 419 [of Tarski, 1983b]) according to which if one treats all non-logical constants as logical constants, the notion of logical consequence coincides with that of material consequence. In a variable domain conception of model it is hard, if not impossible, to make sense of the claim (see also Bays, 2001, pp. 1708–1709). We have seen that Gómez-Torrente (1996a) disagrees with Etchemendy but remarkably he says nothing at all about Tarski's claim (LM). He might be probably tempted to take on this issue the line taken by Ray (1996). (Mancosu, 2006, p. 218)

The line taken by Ray (following Sher, 1991, pp. 45–46) is to say that *either* Tarski was wrong in asserting that logical consequence reduces to material consequence when all constants are taken as logical *or* he did not contemplate domain variation in the test for logical consequence. Since Ray and Sher have argued (with considerations different from mine) that Tarski did contemplate domain variation, they think that Tarski made a mistake when he said that logical consequence reduces to material consequence when all constants are taken as logical. But Ray's and Sher's reason is given by their false belief that Tarski was thinking of conventions entirely similar to the present conventions for the semantics of first-order languages, and in particular that his definition as formulated in (1936b) allowed for domain variation without a corresponding change in the interpretation of a nonlogical predicate. (They don't even consider this hypothesis.) So for them a sentence such as $\exists x \exists y(x \neq y)$ is materially true (when the intended interpretation has more than one individual) and yet is not logically true in the Tarskian sense.

From the point of view of my exegesis, by contrast, there is no need to attribute any error to Tarski, and his claim (LM) is accommodated in a straightforward way. (This may have been the reason why I did not mention it in TLC. It is mentioned, with the accommodation that my exegesis provides, in my dissertation, Gómez-Torrente, 1996b, at p. 52, fn. 24.) $\exists x \exists y(x \neq y)$ is true and logically true in Tarski's sense if its quantifiers range over a domain which contains at least two "logical" individuals, as in the simple theory of types mentioned in Tarski's 1936 paper or in the theories of the 1940 unpublished typescript. If the quantifiers range over a specific domain of objects in a first-order mathematical theory brought into line with the conventions presupposed in Tarski (1936b, 1937a), it is just short for, say, $\exists x \exists y(\mathrm{Zl}(x) \,\&\, \mathrm{Zl}(y) \,\&\, x \neq y)$, which is still true (if the intended extension of 'Zl' has at least two elements) and logically true in Tarski's sense if those conventions are modified exclusively so as to stipulate that 'Zl' is now a logical constant, not subject to reinterpretation. Mancosu's and Bays's criticism is thus seen to be based on conflating my reasons for the variable-domains view with Ray's and Sher's defense of a stronger, and in my opinion inadequate view.

§3. ω-arguments. As we recalled above, Tarski's main motivation for proposing the model-theoretic definition of logical consequence is his idea that certain ω-arguments are logically correct: "... intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$" (Tarski, 1983b, p. 411, 1936b, p. 3), where A and $A_0, A_1, \ldots, A_n, \ldots$ are as follows:

A. Every natural number possesses the given property P.

A_0. 0 possesses the given property P,

A_1. 1 possesses the given property P,

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

. . .

A_n. n possesses the given property P.

. . .

Here "'n' represents any symbol which denotes a natural number in a given (e.g. decimal) number system" (Tarski, 1983b, p. 410, 1936b, p. 2). Yet by Gödel's first incompleteness theorem, if the theory to which A and $A_0, A_1, \ldots, A_n, \ldots$ belong proves some basic arithmetic, there will always be sentences of the form of A that are not provable in the theory even though all the sentences of the form of $A_0, A_1, \ldots, A_n, \ldots$ are so provable. According to Tarski this shows that the notion of derivability cannot be used to define the proper concept of consequence.

As also recalled above, according to Etchemendy Tarski's thought was that ω-arguments in a typical first-order language for elementary arithmetic (arguments with a set of premises of the forms $P0$, $Ps0$, $Pss0$, etc. and conclusion $\forall x Px$) were logically correct. TLC and Sher (1996)[26] pointed out, however, that the textual evidence suggested that Tarski was thinking of ω-arguments in a higher-order language, and by now there seems to be a wide agreement among interpreters that this is the right way of looking at Tarski's assertions. But some issues remain unsettled. First, what was exactly the higher-order form of the ω-arguments that Tarski had in mind? Second, Bays (2001) has claimed, on the basis of his own proposed answer to this question, that Tarski's treatment of ω-arguments shows by itself that he held a fixed-domain view of interpretations and models, in the sense discussed in section 2. We will consider these two issues in turn.

As TLC had done, Bays points to Tarski's reference to his (1933) paper. Here Tarski considers a number of theories formulated in the language of GTC above, showing that they are all ω-incomplete, although by this he means here something slightly different from, if mathematically equivalent to, what he means in his paper on logical consequence. As Bays notes (at p. 1719), the "ω-arguments" here have this intuitive form:

A_0^*. The empty set possesses the given property P,
A_1^*. Sets containing at most one element possess the given property P,
. . .
A_n^*. Sets containing at most n elements possess the given property P,
. . .

A^*. Every finite set possesses the given property P.

(Cf. Tarski, 1933, p. 105; strictly speaking, Bays by mistake reads A_n^* as "Sets containing *exactly* n elements possess the given property P" (cf. Bays, 2001, p. 1719).) To avoid some potential misunderstandings, it is of some importance to make explicit the exact (formalized) form of A^* and $A_0^*, A_1^*, \ldots, A_n^*, \ldots$ In order to do this note first that with third-order quantifiers over classes of sets of individuals we can define a set of first-order individuals to be finite if it belongs to every class of sets that (1) contains the empty set, and (2) if it contains a set a, it contains any set b that adds at most one element to a. Or, equivalently, a set of individuals is finite if it belongs to every class C of sets such that, if a set b (1) is the empty set or (2) adds at most one element to a set a that belongs to C, then b belongs to C. A third-order formula of simple type theory that Tarski uses to express this

[26] Bays (2001), at p. 1719, n. 36, and at p. 1720, n. 41, incorrectly says that Sher (1991) and Ray (1996) also made this claim.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

last idea, with X_1^2 the free variable, is

$$\forall X_1^3(\forall X_2^2\forall X_3^2((\forall X_1^1 \sim X_2^2(X_1^1) \vee X_1^3(X_3^2) \,\&\, \exists X_1^1\forall X_2^1(X_2^2(X_2^1)$$
$$\equiv (X_3^2(X_2^1) \vee X_1^1 = X_2^1)))) \supset X_1^3(X_2^2)) \supset X_1^3(X_1^2)).^{27}$$

Abbreviate this as FINITE(X_1^2). Then the "ω-arguments" considered by Tarski in (1933) all have this form:

A_0^{**}. $\forall X_1^2(\forall X_1^1 \sim X_1^2(X_1^1) \supset P(X_1^2))$,
A_1^{**}. $\forall X_1^2(\exists X_1^1\forall X_2^1(X_1^2(X_2^1) \supset X_1^1 = X_2^1) \supset P(X_1^2))$,
...
A_n^{**}. $\forall X_1^2(\exists X_1^1 \ldots \exists X_n^1\forall X_{n+1}^1(X_1^2(X_{n+1}^1) \supset X_1^1 = X_{n+1}^1 \vee \ldots \vee X_n^1 = X_{n+1}^1)) \supset P(X_1^2))$,
...

A^{**}. $\forall X_1^2(\text{FINITE}(X_1^2) \supset P(X_1^2))$.

(But imagine the formula above in place of 'FINITE(X_1^2)'; 'FINITE(X_1^2)' is an abbreviation, not a defined term.) The property P varies depending on the argument. In the first case considered by Tarski it is given by the formula $\exists X_1^1 \sim X_1^2(X_1^1)$. But Tarski also considers (without constructing it) the formula that would result from a suitable adaptation of the Gödelian construction used in the first incompleteness proof (see below).

Bays claims that the ω-arguments Tarski refers to in (1936b) are arguments of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ (leaving aside the mistake mentioned above). However, this is at odds with the textual evidence. Tarski's relevant passage, quoted in full, says:

> 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 possesses the given property P,
> A_1. 1 possesses the given property P,
>
> And, in general, all particular sentences of the form
>
> A_n. n possesses the given property P,
>
> where 'n' represents 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. (Tarski, 1983b, p. 410, 1936b, p. 2).

Note first that Tarski does not say that the form of argument that he mentions is exactly the form of the argument considered in his (1933) paper, but only that sentences with the forms of $A_0, A_1, \ldots, A_n, \ldots$ are theorems of a theory considered in the paper, while A is not a theorem of that theory (cf. Tarski, 1936b, pp. 2–3, n. 1). Second, note that

[27] Cf. Tarski (1933), p. 105. See also Edwards (2003), p. 53, where this formula is embedded in the sentence called B by Edwards; but beware of a confusing erratum: the third occurrence of 'X_2^2' in B should be one of 'X_3^2'.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

Tarski clearly says that '0', '1',..., 'n',... stand for a "symbol which denotes a natural number in a given (e.g. decimal) number system", but no such symbols appear in $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots$, which only contain connectives, quantifiers and (quantified) variables (and do not use any device of term construction such as a lambda operator). Also, the predicate abbreviated by 'FINITE(X_1^2)' cannot reasonably be the predicate Tarski has in mind when he uses the predicate 'natural number' in A, for in no standard construction of the natural numbers are these represented by all the finite sets; in standard constructions one either picks for each cardinality a certain finite set as representative of that cardinality, or one takes the set of all sets of a cardinality to represent that cardinality.[28] Bays, replying to an imagined objection to his view that "my solution preserves less of the "surface structure" of Tarski's ω-inference example than Etchemendy's solution does", says that "this failure to respect surface structure doesn't bother me very much. Tarski clearly intends his example in (1936b) to be a mere summary of the more detailed example given in (1933)" (Bays, 2001, p. 1721). But for the reasons given, I think that this claim about Tarski's intentions is not clear at all; besides, if Tarski had wanted to give a summary of the exact example of (1933), why would he not have used the informal and perfectly intelligible $A_0^*, A_1^*, \ldots, A_n^*, \ldots / A^*$ (as Bays does)?

Reasons of this sort led me to postulate in TLC an exegesis intermediate between Etchemendy's and Bays's. As Tarski noted, "in the formal language that is in question here [in the theories of (1933)], there do not occur *specific symbols* for the designation of the natural numbers" (Tarski, 1933, p. 105, n. 15, my emphasis). If we attend to his usage of the expression 'specific symbol' elsewhere (see e.g. Tarski, 1937a, p. 12), it's clear that by this he means that there are no *extra-logical primitives* denoting the natural numbers, unlike in the system of Gödel (1931) (quoted in the same note). Since Tarski says explicitly that the ω-arguments he has in mind can be found in the formal language of his theories of (1933) (cf. Tarski, 1983b, p. 410, n. 1), this excludes Etchemendy's exegesis.[29] But Bays's exegesis is excluded also, for example because it's incompatible with Tarski's assertion that $A_0, A_1, \ldots, A_n, \ldots$ *do contain* symbols denoting the natural numbers. A middle course is to postulate that in (1936b) Tarski is thinking of the numerals and the predicate 'natural number' as *logically defined* terms.[30]

---

[28] As Edwards (2003), pp. 53–54, accurately points out (without mentioning Bays), A_0^{**}, $A_1^{**}, \ldots, A_n^{**}, \ldots$ "are not of the forms 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".

[29] It also excludes Jané's suggestion (2006, p. 33) that Tarski (1936b) had in mind at most weak second-order languages in which there is "an individual constant c (for the number zero) and a unary function symbol f (for the successor operation)" as nonlogical symbols, and where the natural numbers are defined by means of the second-order formula (with 'x' the free variable) $\forall X(Xc \,\&\, \forall y(Xy \supset Xfy) \supset Xx)$. (Note also that Jané's suggestion doesn't make sense of Tarski's claim that 'n' may represent "any symbol which denotes a natural number in a given (e.g. decimal) number system", while on the exegesis that the numerals are taken as defined symbols, the claim is accommodated without problems. For more on Jané's view see section 5 below.) Edwards (2003) more cautiously (and I think correctly) claims that Tarski was aware of these languages, and of the fact that arguments of the form $Pc, Pfc, Pffc, \ldots / \forall x(\forall X(Xc \,\&\, \forall y(Xy \supset Xfy) \supset Xx) \supset Px)$ are declared logically correct by his definition (this will be so even if all interpretations share the domain of quantification of the intended interpretation).

[30] I think Edwards (2003) is wrong to hold that Tarski's actual use of an ω-argument from simple type theory reduces the attractiveness of his claims that derivational conceptions do not, and his defined concept does, recover the common notion of consequence. Edwards says this because he thinks that Tarski's example presupposes a "reductive" or logicist conception of arithmetical

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

There are several well-known ways to do this, but presumably Tarski had in mind the most standard, Russellian one (cf. Tarski, 1941, p. 81, ad finem). Let's see how this makes sense of all of Tarski's claims. First, we can introduce by definition each one of the individual numerals (e.g. the *decimal* ones, as used in *Principia Mathematica* (Whitehead and Russell, 1925–1927, vol. I, $_*52$, $_*54$)) as names of classes of classes (the decimal numeral for n denoting the class of classes with n elements):

$$\forall X_1^2(0(X_1^2) \equiv \forall X_1^1 \sim X_1^2(X_1^1)),$$

$$\forall X_1^2(1(X_1^2) \equiv \exists X_1^1(X_1^2(X_1^1) \,\&\, \forall X_2^1(X_1^2(X_2^1) \supset X_1^1 = X_2^1))),$$

$$\ldots$$

$$\forall X_1^2(n(X_1^2) \equiv \exists X_1^1 \ldots \exists X_n^1(X_1^2(X_1^1) \,\&\, \ldots \,\&\, X_1^2(X_n^1) \,\&\, X_1^1 \neq X_2^1 \,\&\, X_1^1 \neq X_3^1 \,\&$$

$$\ldots \,\&\, X_n^1 \neq X_{n-1}^1 \,\&\, \forall X_{n+1}^1(X_1^2(X_{n+1}^1) \supset X_1^1 = X_{n+1}^1 \vee \ldots \vee X_n^1 = X_{n+1}^1))),$$

$$\ldots$$

We can express the relation (in fact function) of successor, X is the successor of Y, among classes of classes, by means of this formula, which we abbreviate as SUCC(X_1^3, X_2^3):

$$\forall X_1^2(X_1^3(X_1^2) \equiv \exists X_2^2(X_2^3(X_2^2) \,\&\, \exists X_1^1(\sim X_2^2(X_1^1) \,\&\, \forall X_2^1(X_1^2(X_2^1)$$

$$\equiv (X_2^2(X_2^1) \vee X_1^1 = X_2^1))))))).$$

Intuitively put, X_1^3 is the successor of X_2^3 if the classes that are in X_1^3 are exactly all the results of adding exactly one individual to some class that belongs to X_2^3. (Note that the successor of the empty set (of the third type) is the empty set itself.) The successor function is then used to define the class of the classes that are going to be the natural numbers. We can define the class of "logical" natural numbers by means of the formula

$$\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \equiv \forall X_1^4(X_1^4(0) \,\&\, \forall X_2^3(X_1^4(X_2^3)$$

$$\supset \exists X_3^3(\text{SUCC}(X_3^3, X_2^3) \,\&\, X_1^4(X_3^3))) \supset X_1^4(X_1^3))).$$

(Again think of 'SUCC(X_3^3,X_2^3)' as an abbreviation.) This intuitively says that X_1^3 is a NATURAL NUMBER if it belongs to every class X_1^4 of classes of classes to which 0 belongs and such that, if a class of classes X_2^3 belongs to X_1^4, then the successor of X_2^3 belongs to X_1^4. If there are infinitely many individuals, the NATURAL NUMBERS are, as desired, classes a_n of classes of individuals, where a_n contains precisely the classes containing exactly n individuals. If there are only finitely many individuals, the NATURAL NUMBERS will be some other, finite class (see below).

notions, while this commitment is not required by the common notion of consequence. But Tarski's examples work independently of any reductive conception that the logicist constructions "are", or even "reduce" the arithmetical notions. As will be seen below, when defined symbols are eliminated from Tarski's arguments, the resulting arguments are still intuitively correct (they are even truth-preserving in all domains of quantification) but are not instances of derivational consequence.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

Let T_ω be as GTC, but without the axiom of infinity and with the infinitely many axioms

$$\alpha_0.\forall X_1^2(\forall X_1^1 \sim X_1^2(X_1^1) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$\alpha_1.\forall X_1^2(\exists X_1^1 \forall X_2^1(X_1^2(X_2^1) \supset X_1^1 = X_2^1) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$\ldots$$
$$\alpha_n.\forall X_1^2(\exists X_1^1 \ldots \exists X_n^1 \forall X_{n+1}^1(X_1^2(X_{n+1}^1) \supset X_1^1 = X_{n+1}^1 \vee \ldots \vee X_n^1 = X_{n+1}^1))$$
$$\supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$\ldots$$

α_n intuitively says that for any set X_1^2 with at most n elements, there is one individual that doesn't belong to X_1^2. Tarski (1933), p. 108, pointed out that although T_ω of course proves all the sentences $\alpha_0, \alpha_1, \ldots, \alpha_n, \ldots$, T_ω does not prove the sentence

$$\alpha_\infty \forall X_1^2(\text{FINITE}(X_1^2) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$

which says that for any finite set X_1^2 there is one individual that doesn't belong to X_1^2, and thus is provably equivalent to more standard formulations of the axiom of infinity in type theory. This gives us a clue as to the sort of sentences that he probably had in mind in Tarski (1936b). For example, every sentence in the series

$$A_0'.\forall X_1^2(0(X_1^2) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$A_1'.\forall X_1^2(1(X_1^2) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$\ldots$$
$$A_n'.\forall X_1^2(n(X_1^2) \supset \exists X_1^1 \sim X_1^2(X_1^1)),$$
$$\ldots$$

is deductively implied by the sentences in the series $\alpha_0, \alpha_1, \ldots, \alpha_n, \ldots$, so all of $A_0', A_1', \ldots, A_n', \ldots$ are also provable in T_ω. However, the sentence

$$A'.\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset \forall X_1^2(X_1^3(X_1^2) \supset \exists X_1^1 \sim X_1^2(X_1^1))),$$

which deductively implies α_∞, is equally not provable in T_ω. It seems most likely that in the opening pages of Tarski (1936b), Tarski had in mind the argument $A_0', A_1', \ldots, A_n', \ldots / A'$, or a closely related one.

(It is equally clear what Tarski probably had in mind when he said that Gödel's theorem showed that we could always find arguments with the properties of $A_0', A_1', \ldots, A_n', \ldots / A'$ in sufficiently powerful logical theories. Assume a Gödel numbering for the language of GTC to be given, and let $\text{PROOF}(X_1^3, X_2^3)$ abbreviate a formula that expresses the relation in which two NATURAL NUMBERS stand when (the number represented by) the first codes a proof of a formula coded by (the number represented by) the second in a theory T based on that language. By Gödel's diagonal construction, there is a fixed point of the formula $\forall X_2^3(\text{NATURAL NUMBER}(X_2^3) \supset\sim \text{PROOF}(X_2^3, X_1^3))$, i.e. a NATURAL NUMBER g such that the sentence coded by (the number represented by) g is deductively equivalent in T with the sentence $\forall X_2^3(\text{NATURAL NUMBER}(X_2^3) \supset\sim \text{PROOF}(X_2^3, g))$ (which thus "says of itself" that it is not provable). Then clearly, if T is minimally powerful,

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

all the sentences

$$A''_0.\sim \text{PROOF}(0, g),$$
$$A''_1.\sim \text{PROOF}(1, g),$$
$$\ldots$$
$$A''_n.\sim \text{PROOF}(n, g),$$
$$\ldots$$

are provable in T. However, the sentence

$$A''.\forall X^3_2(\text{NATURAL NUMBER}(X^3_2) \supset\sim \text{PROOF}(X^3_2, g))$$

will not be provable in T.[31])

Now that we have a good understanding of the sorts of ω-arguments that are at stake, and thus of how the first issue mentioned at the beginning of this section is to be resolved, let's turn to the second issue, that is the question whether, as Bays claims, Tarski's treatment of ω-arguments shows by itself that he held a fixed-domain view of interpretations and models in Tarski(1936b). Bays's reasons for this claim, which he would probably give even if he conceded that what Tarski has in mind is something like $A'_0, A'_1, \ldots, A'_n, \ldots/A'$, are given in this passage:

> Note that (1933) is one of the places where Tarski treats the axiom of infinity as a principle of logic. Nor is this treatment of Infinity surprising: Tarski needs *something like* the axiom of infinity to ensure that his models contain enough elements to formulate the ω-inferences he is interested in. ([footnote:] I.e., to construct all of the finite sets which play the role of numbers in these ω-inferences.) As a result,

[31] What Tarski had in mind when he said that Gödel's theorem showed that we could always find arguments with the properties of $A'_0, A'_1, \ldots, A'_n, \ldots/A'$ but with the form $A^{**}_0, A^{**}_1, \ldots, A^{**}_n, \ldots/A^{**}$ in sufficiently powerful logical theories (cf. Tarski 1933, 109) was, analogously, the following. Assume a Gödel numbering for the language of GTC to be given, and let $\text{PR}(X^2_1, X^3_1)$ abbreviate a formula that expresses the relation in which a finite set and a NATURAL NUMBER stand when the number (cardinality) of the first codes a proof of a formula coded by (the number represented by) the second in a theory T based on that language. Again by Gödel's diagonal construction, there is a NATURAL NUMBER g such that the sentence coded by (the number represented by) g is deductively equivalent in T with the sentence $\forall X^2_1(\text{FINITE}(X^2_1) \supset\sim \text{PR}(X^2_1, g))$; call this sentence A^{***}. Then a minimally powerful T will not prove A^{***} but will prove all of $A^{***}_0, A^{***}_1, \ldots, A^{***}_n, \ldots$:

$$A^{***}_0.\forall X^2_1(\forall X^1_1 \sim X^2_1(X^1_1) \supset\sim \text{PR}(X^2_1, g)),$$
$$A^{***}_1.\forall X^2_1(\exists X^1_1 \forall X^1_2(X^2_1(X^1_2) \supset X^1_1 = X^1_2) \supset\sim \text{PR}(X^2_1, g)),$$
$$\ldots$$
$$A^{***}_n.\forall X^2_1(\exists X^1_1 \ldots \exists X^1_n \forall X^1_{n+1}(X^2_1(X^1_{n+1}) \supset X^1_1$$
$$= X^1_{n+1} \vee \ldots \vee X^1_n = X^1_{n+1})) \supset\sim \text{PR}(X^2_1, g)),$$
$$\ldots$$

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> Tarski's own understanding of his ω-inference example seems to require (at least) an "infinite domain" conception of models. This demand for infinite domains, however, tells strongly against variable-domain readings of Tarski (1936b). After all, part of the point of variable-domain conceptions of models is to allow models to have *arbitrary* cardinality (including arbitrary *finite* cardinality). And this is what Tarski must reject if his analysis of ω-inferences is to be viable. (Bays, 2001, p. 1720)

One might at first sight think that the idea is that the supposition that arguments of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ are logically correct in Tarski's sense is incompatible with the admission of finite interpretations: Tarski must reject models of finite cardinalities "if his analysis of ω-inferences is to be viable", and as an indirect consequence of this must reject a variable-domains view. But this cannot be the point, for even Bays notes that arguments of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ *are* model-theoretically valid given the now standard, variable-domains model theory for type theory (see Bays, 2001, p. 1719). (Below I will stress that also arguments of the form of $A'_0, A'_1, \ldots, A'_n, \ldots / A'$, are model-theoretically valid.) It is then clear that the Tarskian logical correctness of arguments of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ does not imply by itself that the language of GTC cannot have finite interpretations, or even that its intended interpretation is not finite: since every argument of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ is model-theoretically valid, then, no matter what is the domain of quantification of the intended interpretation of GTC, and in particular no matter whether it is infinite or finite, every argument of that form will be truth-preserving in that domain of quantification, and the argument will be logically correct according to Tarski's definition.

The idea seems to be, rather, that two other facts imply that for Tarski all the interpretations of the language of GTC (or T_ω) have infinite domains of quantification, and in effect presumably the same infinite domain: first, the fact that Tarski wants (e.g.) all of $A'_0, A'_1, \ldots, A'_n, \ldots$ and A' to be *principles of logic*; second, the fact that, on Tarski's intended interpretation of GTC, the (defined) numerals should presumably denote pairwise different objects (which they would not in finite interpretations; see below). (Recall that on Bays's reconstruction, Tarski's ω-arguments are of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$, and thus don't contain numerals; but he seems to think that infinity is needed even "to construct all of the finite sets which play the role of numbers in these ω-inferences"; for the reasons above, I find this clarification obscure.) If this is Bays's idea, the answer to it from the point of view of TLC should be obvious after the discussion in subsection 2.2. As we noted, TLC developed the idea that in languages like the language of GTC and T_ω, Tarski was ready to accept sentences like $\exists x \exists y (x \neq y)$ and an axiom of infinity expressed only with logical constants as type-theoretic logical truths about "logical" individuals.[32] For this reason Tarski was ready to accept that there is no variation of the domain of individuals across interpretations of that language, and no conventions are introduced that allow him to vary that domain. (Thus, as already noted in subsection 2.2, there is just one interpretation of the language of GTC, for this language does not have any nonlogical

[32] Thus I again find perplexing a claim about TLC made by Bays, namely that "It is interesting to note that TLC, Ray (1996) and Sher (1991) all discuss the use of higher-order logic in (1933), but then fail to notice the relevance that this has for the question of fixed vrs. variable domains in (1936b)" (Bays, 2001, p. 1720, n. 41). (As noted above, Bays's claim that Ray (1996) and Sher (1991) discuss the use of higher-order logic in Tarski (1933) is actually incorrect.)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

constants.[33]) The variable-domains view in TLC is simply that those conventions are only presupposed by Tarski (1936b) when the domain of quantification is not a "logical" one, as in a typical first-order language for arithmetic. Therefore, the fact that Tarski thought of $A'_0, A'_1, \ldots, A'_n, \ldots$ and A' as principles of logic, and thus that he accepted infinity as a principle of logic, does not prove that he held a fixed-domain view in Tarski (1936b). Similar remarks apply to the claim that the fact that the numerals of GTC should presumably denote pairwise different objects in Tarski's intended interpretation, implies that he held a fixed-domain view. What this fact does is merely to confirm that Tarski was happy to assume a fixed infinite domain in the test for logical truth and consequence for arguments and sentences of the language of GTC.

There is a final important point to stress in connection with domain variability and Tarski's treatment of ω-arguments. Even though Tarski generally stuck to the standard idea that a sentence of GTC is logically true just in case it is true in the domain of the intended interpretation (as noted in subsection 2.2), we saw that he also considered sympathetically the possibility of defining logical consequence for sentences of GTC as preservation of truth in all nonempty domains of quantification, and even said that it seemed "better founded". This might be thought to create a puzzle of a sort. Since Tarski's adoption of the standard idea seems to a great extent to derive from a desire to accommodate normal logicist usage, but not from a conviction that this was the only acceptable usage or even the best one, it would be puzzling if his main criticism of the derivational understanding of logical consequence depended specifically on logicist assumptions. However, this is clearly not so, and the potential puzzle dissolves. The fact that Tarski does not say merely that $A'_0, A'_1, \ldots, A'_n, \ldots$ and A' are to be counted as principles of logic, but also that A' *follows* from $A'_0, A'_1, \ldots, A'_n, \ldots$ (cf. Tarski, 1983b, p. 411, 1936b, p. 3) is essential. If the claim that A' follows from $A'_0, A'_1, \ldots, A'_n, \ldots$ is to be a robust one, one holding both under the standard logicist idea and under the Hilbertian idea viewed sympathetically by Tarski, we should expect A' to follow from $A'_0, A'_1, \ldots, A'_n, \ldots$ even under a definition of logical consequence for logical type-theoretic sentences as preservation of truth in all nonempty domains of quantification. And this is actually what happens.

As already pointed out in TLC (p. 148, n. 7), it is easy to see that the argument $A'_0, A'_1, \ldots, A'_n, \ldots / A'$, and in general the arguments of the form

$$\begin{array}{l} P(0), \\ P(1), \\ \ldots \\ P(n), \\ \ldots \\ \hline \forall X^3_1(\text{NATURAL NUMBER}(X^3_1) \supset P(X^3_1)) \end{array}$$

are all truth-preserving in all nonempty domains of quantification. Of course truth-preservingness holds when the domain of individuals is infinite, but also for any nonempty finite domain D. If D has exactly a finite number m of elements, then the successor of the denotation of the numeral 'm' (the unit set of the set containing those m elements) will be the empty set (of the third type). The successor of the empty set is the empty

[33] See also the remarks in footnote 17.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

set itself, so the NATURAL NUMBERS will form a finite class with the denotations of all numerals for numbers greater than m being identified with the empty set. So clearly $P(0), P(1), \ldots, P(n), \ldots / \forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset P(X_1^3))$ will be truth-preserving in D. Arguments of the form $A_0^{**}, A_1^{**}, \ldots, A_n^{**}, \ldots / A^{**}$ are model-theoretically valid, as we saw that Bays notes,[34] but so are the arguments of the form of $A'_0, A'_1, \ldots, A'_n, \ldots / A'$, which otherwise make much better sense of Tarski's assertions.

Besides, it's also the case that for every consistent deductive theory T that extends GTC minus the axiom of infinity (AxInf) there will be arguments of the form $P(0), P(1), \ldots, P(n), \ldots / \forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset P(X_1^3))$ such that $T+\{P(0), P(1), \ldots, P(n), \ldots\}$ does not prove $\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset P(X_1^3))$. Thus one can show, by means of examples of ω-arguments where the conclusion is not derivable from the premises, that no consistent deductive theory T whose axioms are all true in all domains and whose rules of inference are all truth-preserving in every domain will capture derivationally the notion of logical consequence as truth-preservingness in every domain. As recalled above, by Gödel's theorem there is an argument $Q(0), Q(1), \ldots, Q(n), \ldots / \forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$ such that GTC+$\{Q(0), Q(1), \ldots, Q(n), \ldots\}$ does not prove $\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$; so GTC−AxInf +$\{Q(0), Q(1), \ldots, Q(n), \ldots\}$ does not prove $\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$ either. More generally, if T consistently extends GTC, by Gödel's theorem there is an argument $Q(0), Q(1), \ldots, Q(n), \ldots / \forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$ such that $T+\{Q(0), Q(1), \ldots, Q(n), \ldots\}$ does not prove $\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$; so $T-\text{AxInf}+\{Q(0), Q(1), \ldots, Q(n), \ldots\}$ does not prove $\forall X_1^3(\text{NATURAL NUMBER}(X_1^3) \supset Q(X_1^3))$ either. This strengthens the point that, even though Tarski was happy to make use of a fixed domain in the test for logical truth and consequence in his language of simple type theory, his ω-arguments objection was very robust, and it held even adopting a Hilbertian variable-domains test for logical consequence in that language.

An all-encompassing understanding of Tarski's treatment of ω-arguments should in all probability take account of the fact that the arguments he had in mind, those of the form of $A'_0, A'_1, \ldots, A'_n, \ldots / A'$, are model-theoretically valid, and should incorporate the idea that Tarski must have seen a source of support for his main criticism of derivational accounts in the fact that his paradigm ω-arguments are truth-preserving in all domains. But this can only be consistently claimed if the interpreter acknowledges the fact that Tarski, in his 1936 paper, has in mind both fixed-domain and variable-domains frameworks for the definition of the logical consequence relation.[35]

§4. Modality and consequence. Under most pretheoretic conceptions of logical consequence, if a sentence X follows logically from a set of sentences K, then there is a

[34] Here Bays adds the puzzling comment that the fact that these arguments are model-theoretically valid explains "why Tarski wasn't too worried about [Etchemendy's] problem [that "ω-inferences aren't valid in first-order model-theory"]" (Bays, 2001, p. 1719). This is puzzling because Bays cannot have it both ways. If Tarski is uniformly relying on a fixed-domain conception in his 1936 paper, then he cannot appeal to a (a version of the) variable-domains conception to calm a certain worry that his presentation in the paper might generate.

[35] In the case of a variable-domains framework for the language of GTC, Tarski's model-theoretic apparatus should of course use conventions different from those presupposed in the paper, which are designed for typical axiomatic mathematical theories.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

strong modal connection between the premises in K and the conclusion X: in some sense or senses of "could", it could not be the case that X were false if the sentences in K were true; or, alternatively, in some sense or senses of "must", X must be true if the sentences in K are true. Typically, this sense or senses of "must" are taken to be very strict: the necessity of logical implication is taken to be stronger than other types of modal implication.

For this reason, the first inclination of most people (including myself) is to read a reference to some such strong modal connection in the following passage of Tarski (1936b), which we may call "passage I" for later reference:

> Certain considerations of an intuitive nature will form our starting-point. Consider any class K of sentences and a sentence X which follows from the sentences of this class. From an intuitive standpoint [a] it can never happen that both the class K consists only of true sentences and the sentence X is false. Moreover, since we are concerned here 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, [b1] this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the objects to which the sentence X or the sentences of the class K refer. [b2] The consequence relation cannot be affected by replacing the designations of the objects referred to in these sentences by the designations of any other objects. (Tarski, 1983b, pp. 414–415, 1936b, pp. 6–7)

In particular, most people's first inclination is to read sentence [a] as attributing a strong modal connection to instances of logical consequence. Tarski immediately goes on to say, in what we may call "passage II":

> The two circumstances just indicated, which seem to be very characteristic and essential for the proper concept of consequence, may be jointly expressed in the following statement:
> (F) If, in the sentences of the class K and in the sentence X, the constants—apart from purely logical constants—are replaced by any other constants (like signs being everywhere replaced by like signs), and if we denote the class of sentences thus obtained from K by 'K'', and the sentence obtained from X by 'X'', then the sentence X' must be true provided only that all sentences of the class K' are true. (Tarski, 1983b, p. 415, 1936b, p. 7)[36]

Another passage most relevant to the exegesis of the preceding two is the passage where Tarski introduces his model-theoretic definition (passage III):

> . . . we can define the concept of logical consequence as follows:
> The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is [at the same time] also a model of the sentence X.

[36] In TLC I said that the 'F' that names this condition in both the German text and in the English translation probably stood for "formality". Hitchcock (in Stroińska and Hitchcock, 2002, p. 167) has refuted this conjecture noting that the Polish text names it with a 'W', for "wynikanie" (consequence).

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> It seems to me that everyone who understands the content of the above definition must admit that it agrees quite well with common [linguistic] usage. This becomes still clearer from its various consequences. In particular, it can be proved, on the basis of this definition, that every consequence of true sentences must be true, and also that the consequence relation which holds between given sentences is completely independent of the sense of the extra-logical constants which occur in these sentences. In brief, it can be shown that the condition (F) formulated above is necessary if the sentence X is to follow from the sentences of the class K. (Tarski, 1983b, p. 417, 1936b, p. 9; the words in brackets are in the German, but not in the English version)

Passage III clearly announces that the two properties [a] and [b2] mentioned in passage I can be proved to hold of Tarski's defined notion of logical consequence. Since property [a] is prima facie modal, Etchemendy (1990), pp. 85–86, took Tarski to be announcing a proof that if X follows from K in the defined sense, then necessarily, if all the sentences in K are true, then X is true. Since Etchemendy (1990) had argued on philosophical grounds that this claim is in fact false, and thus that no correct proof of it is possible, he claimed that Tarski had in mind a fallacious proof, and specifically the following: if a sentence X follows from a set of sentences K according to the definition, then there is no model of K which is not also a model of X, but if all the sentences in K were actually true and X actually false, then there would be a model of K in which X was false (namely, the model consisting of the *actual* extensions of the extra-logical constants of K and X). This reasoning, Etchemendy noted, commits a modal fallacy, dubbed by him "Tarski's fallacy": taking P to be 'X follows from K in Tarski's sense', Q to be 'All the sentences in K are true', and R to be 'X is true', what the proof should show is a statement with a modality strengthening the consequent of a conditional,

> If P then necessarily (if Q then R).

But all the mentioned proof establishes is a necessitated conditional,

> Necessarily (if P then (if Q then R)).

Other authors, such as García-Carpintero (1993), pp. 125–127, Hanson (1997), pp. 401, 378, and Sher (1991), pp. 41–42, accepted Etchemendy's modal reading of property [a], but strongly emphasized the fact that Tarski did not give the proof he mentions. They suggested that some such proof might exist, and that Etchemendy's philosophical arguments to the contrary were inadequate. Since then, several philosophical suggestions on the directions that an argument for the modal adequacy of model-theoretic consequence might take have been made, notably by Sher (1996) and Shapiro (1998).

On the other hand, Ray (1996), section 5, and I in TLC, pp. 130, 149 n. 17, and in Gómez-Torrente (1998), made the purely exegetical claim that despite the prima facie impression that everybody shared, in passage III Tarski did not announce a proof for any modal claim. This exegesis has gained the distinguished support of Soames (1999), p. 118, and Jané (2006), p. 39, among others. (It must be emphasized that an acceptance of this exegesis is of course compatible with a pretheoretic conception of logical consequence on which there *is* a strong modal connection between the premises and the conclusion in cases of logical consequence, and thus with a theoretical proposal along the lines of Shapiro or Sher mentioned in the previous paragraph. Some remarks on my own view of the philosophical issue can be found in Gómez-Torrente, 1998/1999.)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

In this section I want to apply again the sorts of considerations that led to this exegesis, in a discussion of a recent proposal (and criticism of earlier views) on how to interpret passage I, recently made by David Hitchcock (in Stroińska & Hitchcock, 2002). Hitchcock interprets property [a] as a modal property, though not as logical or metaphysical necessity but as *temporal* necessity or eternity. He notes, for example, that Tarski uses temporal vocabulary in the enunciation of [a] ("it can never happen"). (And (significantly, as we will note) also in the German version of his definition of logical consequence in passage III: "The sentence *X follows logically* from the sentences of the class *K* if and only if every model of the class *K* is *at the same time* also a model of the sentence *X*" (second emphasis added; the temporal metaphor does not appear in the English translation, as we saw).) Hitchcock suggests that Tarski's uses of modal terms are temporal uses, and that they effect a disguised quantification over times, or as he calls them, "circumstances". In particular, [a] is a temporal property, the property that *K* and *X* have when there is no circumstance in which all the sentences in *K* are true and *X* is false:

> the intuitively grounded condition—that it cannot happen that the implying sentences are true and at the same time the implied sentence is false—could be taken to mean that there are no circumstances in which both the implying sentences are true and the implied sentence is false. (Stroińska & Hitchcock, 2002, p. 170)

Although this reading doesn't see a particularly strong modality in Tarski, there are still many reasons why it cannot be a correct reading, reasons which often parallel the reasons why necessity readings cannot be correct. One basic reason is that in passage III Tarski explicitly says that the claim that his defined notion has properties [a] and [b2] is "in brief" "that the condition (F) formulated above is necessary if the sentence *X* is to follow from the sentences of the class *K*". For Tarski, condition (F) is the conjunction of properties [a] and [b2]. But condition (F) is clearly not a modal condition, either metaphysical or temporal. It is true that (F) contains a 'must', but surely the natural way to read (F) is

> (F′) If, in the sentences of the class *K* and in the sentence *X*, the constants—apart from purely logical constants—are replaced by any constants (like signs being everywhere replaced by like signs), then for every class *K′* of sentences thus obtained from *K*, and for every sentence *X′* thus obtained from *X*, the sentence *X′* is true provided only that all sentences of the class *K′* are true.

As far as I know, all scholars who have dealt with Tarski (1936b) have read (F) as (F′),[37] including Etchemendy and, as far as I can tell, Hitchcock. As noted in Gómez-Torrente (1998), everybody has tacitly taken the use of 'must' in (F) as a simple stress of a restricted universal generalization, one that says that a certain property (truth-preservation) holds of all argument instances of a certain kind (the suitable substitution instances of argument *K*/*X*); nobody has taken it to signal necessity or eternity. It's as when we say "If someone kills someone else, then necessarily he goes to jail", simply to stress the universal generalization that in all of an implicitly restricted series of cases in which someone kills someone else, he goes to jail. Nobody has taken this to mean either that in all possible worlds, if

[37] Perhaps with the recent exception of García-Carpintero (2003), who seems tempted to reject this reading and read (F) modally as well.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

someone kills someone else he goes to jail, or that at all times, if someone kills someone else, he goes to jail (there were killings well before there were jails).

But *could* the 'must' in (F) be read metaphysically or temporally, even if no one has in fact so read it? There are several reasons why it could not. One is simply that Tarski says, almost immediately after passage II, that if (F) were not only a necessary but also a sufficient condition for consequence, then

> the problem of formulating an adequate definition of the concept of consequence would be solved affirmatively. The only difficulty would be connected with the term 'true' which occurs in the condition (F). But this term can be exactly and adequately defined in semantics. (Tarski, 1983b, p. 415, 1936b, p. 7)

Tarski would not say this if he thought that (F) contained 'necessarily' or 'eternally'. For, even if it had not been entirely out of the question that such concepts could be given an "exact and adequate" definition from his point of view at the time, he surely would not have said that there was *no difficulty* with defining them "exactly an adequately". There was simply no exact and adequate definition in semantics of the notions of "truth at a world" or "truth at a time" (or "truth at a circumstance") at the time Tarski was writing. Relatedly, note that, as Hitchcock himself says, Tarski uses temporal terms even in his enunciation of the model-theoretic definition of consequence. Since it is evident that Tarski thinks that the definition can be fully given in terms of "scientific semantics", and no metaphysical or temporal terms appear in his contemporary metalanguages for "scientific semantics", the appearance of temporal terms in the definition can only be taken as a usual metaphor among mathematicians. (As in "A prime number is *never* divisible by any other numbers than 1 and itself".) Thus his use of temporal terms in related contexts is in all likelihood a metaphorical use as well.

Perhaps more decisively, under Hitchcock's reading Tarski's claims about the relation between his definition and condition (F) would imply a trivialization of the notion of truth at all times. On this reading, (F) is to be understood as

> (F_T) If, in the sentences of the class K and in the sentence X, the constants—apart from purely logical constants—are replaced by any constants (like signs being everywhere replaced by like signs), then for every class K' of sentences thus obtained from K, and for every sentence X' thus obtained from X, for all times t (the sentence X' is true at t provided only that all sentences of the class K' are true at t).

Surely, under any theoretically worthy conception of the notion of "truth at a time", there cannot be a collapse of the distinction between truth *simpliciter* (or present truth) and truth at all times (or eternal truth)—just as under any theoretically worthy conception of the notion of "truth at a world", there cannot be a collapse of the distinction between truth *simpliciter* (or actual truth) and truth at all worlds (or necessary truth). But Tarski's view would imply such a collapse under Hitchcock's reading. Recall that Tarski says that

> 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. The sentence X would in this case follow from

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> the class K of sentences if either X were true or at least one sentence of the class K were false. (Tarski, 1936b, p. 419)

So, for example, if all terms are taken to be logical and K is the empty set, X is true only if X is true in all interpretations. But since Tarski thought (as noted above) that it can be proved that X is true in all interpretations only if the argument K/X satisfies condition (F), it follows under Hitchcock's reading that Tarski thought that X is true in all interpretations only if the argument K/X satisfies condition (F_T). Or in other words, that X is true in all interpretations only if for all replacements of the nonlogical constants, X is true at all times; or, eliminating the vacuous quantification, that X is true in all interpretations only if X is true at all times. It follows that Tarski was committed to the claim that X is true only if X is true at all times. Since this would trivialize the notion of truth at all times, it is strong indication that Tarski did not think of (F) as a temporal condition. Of course, this line of reasoning can be applied as an additional consideration to show that Tarski did not read (F) as a metaphysical modal condition either. But it cannot be applied against the natural (F') reading, for in this case the only implication is the triviality that, if all terms are taken as logical and X is true, then X (the only sentence which results from X by replacement of its logical terms) is true.

One of Hitchcock's motivations for his exegesis is his dissatisfaction with my reading of the passages at stake:

> Gómez-Torrente (1998) construes the modal expressions in the intuitively based condition and in condition (F) as merely signs of generality, which could be eliminated without loss of meaning; this interpretation however makes the condition of necessary truth-preservation [a] duplicate the formality condition [b2], a duplication which it would be hard to imagine Tarski failing to note. (Stroińska & Hitchcock, 2002, p. 169)

Hitchcock has in mind the following passage from Gómez-Torrente (1998):

> the 'can never' in [a] may be there simply to signal generality, just as 'must' clearly does in (F); more specifically, with [a] Tarski is clearly just saying that all arguments of the same form as a correct argument $< K, X >$ are truth-preserving. Put together with [b2], which gives more concrete content to the notion of "form", [a] results in condition (F). (p. 231)

He is complaining that on my reading [b2] is stronger than [a], and this is implausible, since Tarski would seem to think of them as independent, even if he doesn't say so explicitly. I agree with Hitchcock that this puts some pressure on my claim that "with [a] Tarski is clearly just saying that all arguments of the same form as a correct argument $< K, X >$ are truth-preserving". I admit that Hitchcock's consideration indicates that this is not so clear. Yet (F) is *clearly not* a modal condition, so [a] cannot be a modal condition either. Is there a way to interpret [a] that accommodates all these considerations? Yes, there is—and it is remarkably simple. Hitchcock fails to point out (and I failed to note) that if we wish to read the 'can never' in [a] as signaling mere generality (as we are all but required to do), we are not forced to read this generality as quantifying over all arguments of the same form as K/X. It may quantify over *all* arguments that are correct, *simpliciter*. When Tarski says that "it can never happen that both the class K consists only of true sentences and the sentence

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

X is false", he is probably simply saying that if K/X is *any correct argument*, it is not the case that the class K consists only of true sentences and the sentence X is false. With [b2] he means that all the suitable substitution instances of K/X with constants that name *other objects* than those named with the constants in K/X are equally truth-preserving.[38] On this reading, [b2] is not stronger than [a]. Put together, [a] and [b2] result in the condition that all the suitable replacement instances of K/X with any constants are truth-preserving, that is condition (F) (understood, of course, as (F′)).[39]

Finally, it must be stressed once more that a nonmodal, "generality" reading makes full sense of Tarski's announcement in passage III of the proof that Etchemendy took to be modal and fallacious ("it can be shown that the condition (F) formulated above is necessary if the sentence X is to follow from the sentences of the class K"). On this reading, the proof Tarski is announcing is a proof that if X is true in all the models of K, then the argument K/X satisfies condition (F′). So presumably the proof Tarski had in mind was this: suppose that there is no model of K which is not a model of X; then there is no substitution instance K'/X' of K/X such that the sentences in K' are true and X' is false; for if there was one such, it would readily provide a sequence of objects (the extensions of the extra-logical

[38] Tarski presents [b2] as a more precise or manageable version of [b1] above. While [b1] is an epistemic claim of some sort, [b2], which is the version codified in (F), is not epistemic in any sense. On the basis of [b1], Hanson (1997, p. 378) has attributed to Tarski the strong modal (epistemic) claim that "any knowledge we may have about instances in which the relation [of logical consequence] does or does not hold should be possible to have a priori". But I think it's reasonably clear that [b1] has no such implication about apriority. [b1] seems to say simply that if an argument K/X is logically correct, then no empirical knowledge of the objects named by extra-logical constants we could put in place of the extra-logical constants of K/X would lead to knowledge that the resulting argument K'/X' is not logically correct. This is compatible with any knowledge that K/X or K'/X' are logically correct being reachable only a posteriori.

[39] As we saw, (F) speaks of replacement instances with "any other [non-logical] constants". This could conceivably mean "any non-logical constants of expression types strictly different from those of the non-logical constants appearing in K/X" (in which case it would still include replacements with different constants, if they exist, that name the same objects as the non-logical constant types in the original argument K/X; but such constants may not exist). But I don't think that's what's meant, so in (F′) I have deleted the word "other" for the sake of clarity, taking the quantifier "any" to range over all nonlogical constant expression types of the language at stake. The inclusion of the word "other" in the original (F) is either not intended to indicate a restriction to replacements with constants of strictly different expression types from those of the constants in K/X or, if it is, it's probably a slip on Tarski's part. Note that if "any other constants" means "constants of any types strictly different from those of the non-logical constants appearing in K/X", then condition (F), *under any reading*, is clearly weaker than Tarski intends it to be elsewhere. He considers the possibility of offering (F) as a characterization of consequence, giving not just a necessary but a sufficient condition, and notes that this would only do if "the designations of all possible objects occurred in the language in question" (Tarski, 1983b, p. 416, 1936b, p. 7). But even if "the designations of all possible objects occurred in the language in question", (F′) would still be clearly weak if "any other constants" meant "constants of any strictly different types": (F′) might leave out of consideration the potential counterexample to the truth-preservingness of K/X constituted by the sequence of denotations of the original nonlogical constants of K/X. The same can be said of (F_T): even if "the designations of all possible objects occurred in the language in question", (F_T) (with if "any other constants" meaning "constants of any different types") might leave out of consideration the potential counterexamples to the eternal truth-preservingness of K/X constituted by the sequences of denotations, with respect to the different times t, of the original nonlogical constants of K/X. Mutatis mutandis for a reading of (F) in terms of metaphysical modality.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

constants of K'/X') which would constitute a model of K that would not be a model of X.[40]

§5. Tarski's pretheoretic ideas about logical consequence. Jané (2006) has rightly stressed the lack of extended discussion of Tarski's pretheoretic ideas about logical consequence in the exegetical literature, and the brief remarks on the topic in TLC cannot aspire to constitute an exception. Jané provides such extended discussion in his recent paper. However, I think that this valuable discussion clearly oversimplifies Tarski's pretheoretic conception and misses some of the basic facts about his motivations for proposing his theory of logical consequence the way he did. In this last section I will take Jané's exegesis as a convenient foil for a more accurate, but still rough, sketch of Tarski's pretheoretic views.

Before I start, it may be good to make it clear that I'm not using the term 'pretheoretic' in the same sense as Jané. As we are about to see, he claims that the informal concept of consequence that Tarski wants to characterize is simply the concept, preexisting in mathematical practice, of truth in all informal interpretations of an axiomatic theory. He denies that this concept is pretheoretic, understanding this term to mean "previous to any theoretical activity whatever", and he attributes the idea that Tarski's informal concept of logical consequence was pretheoretic in this sense to TLC and other works. But there could hardly be a "pretheoretic" concept of *logical* consequence in this sense (surely

---

[40] I first suggested that this is the proof Tarski had in mind in Gómez-Torrente (1996a), p. 130; essentially the same suggestion was made by Ray (1996), section 5. As noted in Gómez-Torrente (2000), additional evidence can be gathered from a passage from Tarski (1937b), where Tarski again passes quickly from the existence of a set of true axioms to the existence of a model of the set. After noting, as in Tarski (1936b), that a set of axioms can be transformed into a set of sentential functions, he says:

> When one considers some objects or others, one may examine the question whether they satisfy the system of axioms thus transformed, that is, whether the names of these objects, put in place of the variables, turn these axioms into true propositions; if it turns out that this is so, we will say that the said objects form a *model of the system of axioms* just considered. One such model is constituted, for example, by the objects designated by the primitive terms. (Tarski, 1937b, p. 102).

In Gómez-Torrente (1996b), p. 50, n. 21, and at greater length in Gómez-Torrente (2000), I pointed out that the proof does not work straightforwardly for the usual model-theoretic notion of logical consequence. On the latter, interpretations are made up of sets of the cumulative hierarchy. So for all we know, some arguments K/X may be instances of logical consequence in the usual model-theoretic sense, and yet may have proper-class counterexamples. If the distinguished objects of one such counterexample have names in the language of K/X, then there will be a substitution instance K'/X' of K/X such that all the sentences in K' are true and X' is false. (Although it is unlikely that arguments with the properties of K/X exist (for discussion, see Gómez-Torrente, 1996b, chapter 3, 1998/1999, section 3), the proof we are attributing to Tarski does not straightforwardly exclude this possibility.) In Gómez-Torrente (2000) I proposed that around this time Tarski probably took the metatheory in which one defines logical consequence for a particular language to include in its range of quantification the objects that constitute the standard interpretation of the language, and that he didn't insist (as later practice would make common) on using a standard set theory as a universal metatheory. I conjecture that one of the "open questions" (Tarski, 1983b, p. 418, 1936b, p. 10) that he thought were left by his definition of logical consequence had to do with the need to clarify the scope of the metatheory or metatheories appropriate for the application of the definition.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

any understanding of the notion of logicality presupposes some degree of exposure to theoretical reflection), and I have certainly never claimed that there is or that Tarski thought that there was. In this context what I have always meant, and will continue to mean, is "previous to the theoretical activity of mathematical explication or characterization".

Jané's fundamental conclusions are that "the intended notion of consequence at play in axiomatics since the end of the 19th century is precisely the one that Tarski's definition succeeded in characterizing", and that "Tarski's aim was in fact to characterize this notion of consequence (. . . this particular notion was his *common* concept of consequence)" (Jané, 2006, p. 30). "The intended notion of consequence at play in axiomatics" was a close informal analogue of the notion of consequence defined by Tarski, which as Jané has shown in detail, was (often implicitly) in the mind of many late nineteenth-century and early twentieth-century mathematicians. Given an axiom system $\mathbf{A}(a, b, c \ldots)$ framed in a language with a set of terms $a, b, c \ldots$ specific to the theory (e.g., the distinctive primitives of Euclidean geometry, or of arithmetic, etc.), the idea was that a sentence S is not a consequence of $\mathbf{A}(a, b, c)$ just in case there is a way of interpreting the specific symbols (regardless of their original intended meaning) that in some intuitive sense verifies $\mathbf{A}(a, b, c)$ but falsifies S; or, equivalently, that a sentence S is a consequence of $\mathbf{A}(a, b, c)$ just in case there is no way of interpreting the specific symbols that verifies $\mathbf{A}(a, b, c)$ but falsifies S. In Tarski, the distinctive specific symbols of an axiom system become the so-called extra-logical constants, the defined notion of satisfaction is a precise version of the intuitive notion of "verification" at play in axiomatics, and "ways of interpreting" become interpretations (and models). Thus, according to Jané, Tarski's defined notion of logical consequence as preservation of satisfaction in all models of the premises mirrors without remainder the informal notion of preservation of "verification" in all "ways of interpreting" the specific symbols that "verify" the axioms; it only differs from the latter notion in that it is mathematically precise.

While there is little doubt that the informal ideas on models of axiom systems that developed in late nineteenth-century to early twentieth-century mathematical axiomatics are one of the basic ingredients of Tarski's pretheoretic ideas about logical consequence, I also think that there can be little doubt that these ideas do not reduce to an identification of the relevant pretheoretic concept of logical consequence with the notion of consequence implicit in the axiomatic tradition. Let me start by showing that the texts that Jané takes as positive evidence that Tarski's "common concept" of logical consequence was identical with the informal notion of consequence in axiomatics in fact provide no support for the strong claim of identity. After that I will show that Jané's identity claim is manifestly at odds with many passages from Tarski. The discussion of the positive and negative evidence relative to Jané's claim will help me paint a more balanced picture of Tarski's pretheoretic conception, and especially of those of its aspects that go beyond the informal notion of late nineteenth-century mathematical axiomatics.

According to Jané, "under the assumption that Tarski's common concept of consequence is the one at play in axiomatic theories (. . .) we understand (. . .) why he identifies logical consequence with formal consequence as a matter of course, without attempting to justify it at all" (Jané, 2006, p. 30). As we saw, Tarski says that one "consideration of an intuitive nature" about logical consequence is that it is formal, in the sense that "the consequence relation cannot be affected by replacing the designations of the objects referred to in [the sentences of a correct argument] by the designations of any other objects" (Tarski, 1983b, p. 415, 1936b, pp. 6–7). It is clear that the notion of consequence in axiomatics also has this property, so as Jané says, if Tarski's "common concept" is this notion, then no justification

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

for the claim of formality is needed. But this in no way indicates that Tarski is only thinking of the notion of axiomatics. The idea that logical consequence is formal, in precisely Tarski's sense, is an idea that can be traced back to Aristotle, and that receives explicit formulations in many authors starting with Alexander of Aphrodisias. It is in fact one of the basic traditional traits of the logical consequence relation, which of course precedes nineteenth-century ideas about axiomatic mathematical theories and models. Thus there was no reason why Tarski should have justified that logical consequence is formal, provided merely that he did not have in mind some eccentric pretheoretic ideas about logical consequence or in particular about the "common concept" of consequence.[41]

Jané claims that passage III, quoted in section 4 above, and in particular the sentence "it seems to me that everyone who understands the content of the above definition must admit that it agrees quite well with common usage" (Tarski, 1983b, p. 417) is illuminated by his identity claim. According to Jané,

> under our assumption, these words can be taken at face value, for whoever understands the notions involved in Tarski's definition (and is versed in axiomatics, as contemporary logicians certainly were), recognizes that the definition yields a precise, rigorous version of the consequence relation at play in axiomatics. (Jané, 2006, p. 31)

I find this claim puzzling, for surely one would not expect to hear Tarski say merely that his definition agrees "quite well" (*recht gut*) with common usage if Jané's identity claim held; Tarski would say "the definition *fully* agrees with the common notion, of which it is just a mathematically precise version", or something like that. It is important to point out that Jané's reading is also strongly disconfirmed by the German text, which, as we saw in section 4, literally says "agrees quite well with common *linguistic* usage" (Tarski, 1936b, p. 9, my emphasis), thus throwing doubts on the idea that Tarski has in mind an exclusively mathematical notion; and also by the Polish text, which says "[the definition] captures *many* intuitions manifested in the *everyday* usage of the concept of following" (Tarski, 1936a, p. 186, my emphases). Furthermore, if Tarski had in mind what Jané attributes to him, he would not say that the fact that the definition "agrees quite well with common usage" "becomes still clearer from its various consequences. In particular, it can be proved (. . .) that the condition (F) formulated above is necessary if the sentence X is to follow from the sentences of the class K" (Tarski, 1983b, p. 417, 1936b, p. 9). The fact that (F) is a necessary condition of arguments satisfying Tarski's defined notion of consequence cannot make "clearer" something that should be perfectly obvious, if Jané's reading were right.

Finally, Jané claims that his reading dissolves the following alleged puzzle:

> [Tarski's] choice of an instance of consequence that is not accounted for by the formalized concept was a particular ω-argument. How does he know that this particular argument will turn out to be correct under his definition of consequence, if its correctness should depend on a distinction between logical and extra-logical terms that he has not given? Moreover, the terms that have to be taken as logical and extra-logical in order for that argument to be correct under his definition of

[41] Of course, that logical consequence in Tarski's defined sense is formal is evident from the Tarskian definition.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

> logical consequence are not among the uncontroversial ones that belong to propositional or first-order logic. So, we must admit that he has not shown us that his definition fares better, as regards this particular argument (which he took as a touchstone of consequence), than the formalized concept that he dismissed. (Jané, 2006, p. 32)

According to Jané, the puzzle dissolves

> if Tarski's aim was, as I claim, to define the concept of consequence at work in axiomatic considerations. For the crucial distinction relating to the primitive terms of an axiomatic theory is not between logical and extra-logical in some absolute sense, but between specific and non-specific. (. . .) It is because it deals with the logical/extra-logical distinction in this immanent, non-committal way that Tarski can ensure without checking each particular case that his definition agrees with common (that is to say, axiomatic) usage. In particular, he can ensure that his definition will certify as correct those ω-arguments that are deemed correct by ordinary mathematical considerations. This only requires that the logical (i.e., non-specific) ingredients of the language be sufficient to mimic Dedekind's definition of the set of natural numbers in terms of zero and the successor operation. Second-order quantification or some weak applied set theory is enough for this. (Jané, 2006, pp. 32–33)

In other words: it is puzzling that Tarski should have chosen, as his example of consequence that is not an instance of derivational consequence, an argument whose validity requires embracing the view that there are logical expressions not belonging to first-order languages, for these are not uncontroversially logical, and thus not logical in an "absolute" sense. But on the axiomatic conception, what expressions are specific (extra-logical) and nonspecific (logical) is a matter relative to the theory, or even the context of investigation of a theory. So, provided some theories have second-order quantifiers treated as logical in some contexts (and some way to define zero and the successor operation), his example will work, without the need to suppose that second-order quantifiers are logical across the board.

However, it is again clear that the puzzle does not constitute evidence for Jané's restrictive reading. Even if we concede (as seems reasonable) that for Tarski higher-order quantifiers are not uncontroversially clear cases of logical expressions in all contexts of investigation, this is consistent with Tarski having had in mind a "common concept" of logical consequence not exactly identical with the axiomatic one. For surely, even if Tarski's pretheoretic ideas included elements from other intuitions or traditions, and naturally conceding that the logical/nonlogical distinction has not been made in any precise way nor in any absolute way, his example will appear convincing to someone sharing those ideas provided merely that they include the idea that higher-order quantifiers are logical *in the context of investigation characteristic of the example*. Someone may acknowledge the lack of an uncontroversial characterization of logical constanthood and still rely on the fact that higher-order quantifiers are uncontroversially treated as logical in some contexts, even if these contexts are not contexts that presuppose the axiomatic concept. In fact, the ω-arguments that we *know* Tarski had in mind point to this, and directly cast doubt on Jané's identity claim. We saw in section 3 that, regardless of our preferred view about the concrete ω-arguments that Tarski had in mind, they were, uncontroversially, *arguments formulated in GTC*, that is in the simple type theory that by Tarski's own lights (in Tarski, 1933,

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

p. 105, n. 15), as we also saw, does *not* have nonlogical constants, and thus does not have specific terms either. So it is evident that the example Tarski had in mind belongs to a language that is simply outside the scope of the axiomatic tradition. Tarski's pretheoretic ideas about logical consequence, and presumably their associated pretheoretic concept, thus included aspects that the axiomatic conception did not include.

Jané's alleged positive evidence for his identity claim is, I think, poor. But the negative evidence against the claim is, I think, as rich as the positive evidence is poor. To begin with a point indirectly relevant to the issue we just assessed, which I think clearly disproves the identity claim, recall that Tarski leaves it as an *open question* whether the logical/nonlogical distinction can be given an absolute sense or must be understood as a relative distinction (cf. Tarski, 1983b, p. 420, 1936b, p. 11). Obviously, by Jané's own lights, the axiomatic notion presupposed the *relativity* of the distinction, so if Tarski had thought that his definition of logical consequence fully agreed with the axiomatic notion, he would not have left the question whether the logical/nonlogical distinction can be given an absolute sense or must be understood as a relative distinction as an open problem; he would have said that the logical/nonlogical distinction is a relative distinction, without qualification.

In the second place, Tarski often refers to the common concept of consequence as a concept of "everyday language", for example when he says that

> with respect to the clarity of its content the common concept of consequence is in no way superior to other concepts of everyday language. Its extension is not sharply bounded and its [linguistic] usage fluctuates. Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. We must reconcile ourselves from the start to the fact that every precise definition of this concept will show arbitrary features to a greater or less degree. (Tarski, 1983b, p. 409, 1936b, p. 1; the word in brackets is not in the English version)

Surely the concept of consequence at play in axiomatics is not a concept of everyday language. Besides, it is not a particularly vague or imprecise concept, unlike the common concept of consequence as understood by Tarski; certainly, if we accept Jané's claim that the concept of consequence at play in axiomatics is clearly coextensional with the concept of consequence defined by Tarski it follows, given that the latter is not a vague concept, that the former is not vague either. Jané attempts to dismiss evidence of this sort by appeal to the immediately following passage:

> Even until recently many logicians believed that they had succeeded, by means of a relatively meagre stock of concepts, in grasping almost exactly the common content of the concept of consequence, or rather in defining a new concept which coincided in extent with the common one. (Tarski, 1983b, p. 409, 1936b, p. 1; the English version has "the content of the common concept" instead of "the common content of the concept")

Jané (2006), p. 2, points out that it would be odd for Tarski to say that logicians believed that the derivational concept "coincided in extent with the common one" if the common concept were a vague concept, for a vague concept cannot coincide in extent with a precise one, and logicians should have been aware of this basic point. But note that the full sentence says that what they believed was that they had succeeded "in grasping *almost* exactly the

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

common content of the concept of consequence, or rather in defining a new concept which coincided in extent with the common one" (my emphasis). The phrase after the comma is obviously a clarification of the metaphorical description of the phrase before the comma. In the phrase before the comma Tarski speaks simply of the logicians as "grasping" the common concept, while in the sentence after the comma he speaks of them as "defining a new concept which coincided in extent with" the common concept. But if one phrase is just a less metaphorical version of the other, surely the qualification "almost" must apply to both. This becomes even clearer if we further note that, if Tarski had thought that what the logicians believed was that the derivational concept coincided exactly in extent with the common concept, he would not have said that they believed that they had "almost exactly" grasped the common concept; he would have said that they believed that they had exactly grasped the common concept. Tarski's omission of a qualification like "almost exactly" in the second phrase is probably to be explained by an implicit (if not always appropriate) inclination to avoid redundancy.

More decisively, there are other places where Tarski uses qualifications similar to "almost exactly" to describe the relationship between his defined precise concept of consequence and the common concept, thus implying that while the common concept is vague, his own defined concept at least includes in its extension the clear cases and leaves out the clear non-cases. For example: "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 …" (Tarski, 1983b, p. 413; 1936b, pp. 4–5, my emphasis; the Polish says "In order to obtain the proper concept of following, *essentially* close to the *everyday* concept …" (Tarski 1936a, 181, my emphases)); or in the passage already quoted above: "it seems to me that everyone who understands the content of the above definition must admit that it agrees *quite well* with common *linguistic* usage" (Tarski, 1936b, p. 9, my emphases); or, especially significantly, when Tarski begins to assess the possibility that one could obtain an appropriate derivational concept simply by supplementing the old rules of inference with new rules: "the supposition suggests itself that [by adding new rules of inference] (…) we would succeed finally in capturing the '*essential*' content of the concept of following, which has by no means been exhausted by the rules used until now" (Tarski, 1936a, p. 181, my emphasis); this clearly presupposes that the old derivational concept was also intended to capture the nonvague, or essential core of a vague notion. Most of these passages are simply not mentioned by Jané, but referring to the phrase "close in essentials" in the first of the passages, he only manages to say that "it is hard to discern what this exactly means" (2006, p. 8); I submit that it means exactly what it seems to mean.

There is also evidence against Jané's specific claims about how Tarski's view of ω-arguments supports the claims against the vagueness of the "common concept of consequence". After introducing the argument form $A_0, A_1, \ldots, A_n, \ldots/A$ discussed in section 3 and pointing out that with the usual rules of inference it is possible to prove $A_0, A_1, \ldots, A_n, \ldots$ but not A, Tarski says:

> This fact seems to me to speak for itself. It 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 $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true. (Tarski, 1983b, pp. 410–411, 1936b, pp. 2–3)

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

According to Jané, if Tarski was thinking of the common concept as a vague one, he would not say that "intuitively it seems certain" that A follows from the $A_0, A_1, \ldots, A_n, \ldots$. "He should say, for example, that ω-inferences are so central to our idea of consequence that any account that leaves them aside is unacceptable, that they belong to the core of consequence" (Jané, 2006, p. 3). Jané is of course quite right that if Tarski was thinking of the common concept as a vague one, then his argument against the derivational conception must be based on the idea that ω-arguments are clear cases of the common concept. But I find strange his claim that Tarski should then say more than that "intuitively it seems certain" that A follows from $A_0, A_1, \ldots, A_n, \ldots$. To me this pronouncement definitely conveys the idea that ω-arguments are clear cases of consequence. Even more decisively, in the Polish version, when Tarski strengthens his point by appeal to Gödel's theorem he says, again about ω-arguments and without leaving any trace of doubt, that "we shall be able to construct sentences which *follow in the everyday sense* from the theorems of the deductive theory under consideration, but which cannot be proven in this theory on the basis of the accepted rules" (Tarski, 1936a, p. 181, my emphasis). Indisputably, Tarski would not say this if he thought merely (as Jané implies) that ω-arguments are borderline or otherwise blurry cases of the obviously vague everyday concept (and clear cases only in some contexts that presuppose the axiomatic concept).

Perhaps part of the difficulty leading to Jané's view here is that it may sound odd at first sight to speak of arguments which require a relatively technical formulation, such as the ω-arguments Tarski has in mind, as instances of consequence in the everyday sense. But the fact that an argument contains technical terminology, or even technical logical terminology, is no obstacle to its being an argument that is valid on account of the nature of the everyday concept. In the same way, "f is a function" implies "f is a single-valued correspondence" on account of the nature of the nontechnical concept of (nonlogical) implication (it is an implication by virtue of a definition)—and it does so regardless of what are the reactions of a mathematically untrained or unreflective man-in-the-street about whether there is implication or not between the two sentences.

These facts, and other facts we have reviewed in the course of this paper, suggest a certain picture of Tarski's pretheoretic ideas about logical consequence and the "common concept of consequence" built-in inside them, a picture that I will attempt to describe very briefly and roughly by way of conclusion.

Tarski's "common concept of consequence" is, as he repeatedly indicates leaving little room for doubt, the vague and imprecise common concept of consequence preexisting in ordinary nonmathematical linguistic practice. It is the intuitions about this concept that are the ultimate source for our developed, progressively more and more theoretical, views on what follows, or should count as following, from what. Tarski speaks of the theoreticians, thus presumably of the proponents of derivational calculi, who attempted to grasp the content of the common concept, as having made "efforts . . . to adhere to the common usage of the language of everyday life" (Tarski, 1983b, p. 409, 1936b, p. 1). Being an imprecise concept not subjected to theoretical refinement, the common concept of consequence is inevitably associated with "vague, sometimes contradictory, tendencies" (Tarski, 1983b, p. 409, 1936b, p. 1), which lead the theoretician to make definite, to some extent arbitrary, decisions when she defines technical surrogates. One can only conjecture what obscurities Tarski had in mind, but he may have thought, for example, of the traditional contradictory intuitions about the conditional; here one type of intuition leads to an acceptance of sentences of the form 'If A, then, if B then A' as valid, while another sees these sentences as not generally valid. The classical derivational theoretician abstracts from the second type

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

of intuitions, chooses a truth-functional meaning for the conditional sign, and constructs accordingly her derivational calculus, on which sentences of the form $A \supset (B \supset A)$ come out derivable. (Other derivational theoreticians were of course free to make, and did make, other choices, as in Lewis, 1918; Lewis & Langford, 1932.)

An idea grounded in very basic intuitions about the concept of consequence is the idea mentioned by Tarski (as noted above, in section 4), that in correct arguments it is never the case that the premises are true and the conclusion false. This idea is certainly previous to mathematical ideas or treatments of logical notions or properties, and yet provides a fundamental criterion for logical correctness, so it is not surprising that Tarski singles it out as primary. (In the Polish version Tarski speaks of this idea as arising "from the point of view of *everyday intuitions*" (Tarski, 1936a, p. 183, my emphasis).) An idea that is slightly less fundamental, but is still traceable to basic conceptual intuitions, is the idea also mentioned by Tarski (and also noted in section 4) that logical consequence is formal, that is that all arguments with the same form as a logically correct argument are truth-preserving too. As we saw, Tarski speaks of *both* of these ideas as arising from "considerations of an *intuitive* nature" (Tarski, 1983b, p. 414, 1936b, p. 6, my emphasis). The second idea was part of the logical tradition ever since antiquity, and it can be traced to basic intuitions that in a correct argument some set of expressions in the argument is responsible for its correctness, and that when one changes the parts that are not directly responsible for correctness, one still gets a truth-preserving argument. The idea of "logicality", that some expressions have some distinctive traits that make them especially worthy of study by a theory of correct reasoning, has of course a higher degree of theoreticality, but again was part of ancient logical tradition, and can be traced to a fairly basic, certainly prematheматical, intuition that some expressions are more significant for general reasoning than others.

Presumably a further basic, prematheматical intuition that Tarski appeals to, is the idea that the correctness of an argument is independent of the richness of the class of expressions that one could substitute for the expressions that are not directly responsible for the correctness of the argument (cf. Tarski, 1983b, pp. 415–416, 1936b, pp. 7–8). The intuition that there may be unnamed counterexamples to a certain argumentative form seems again quite fundamental and not greatly theory-laden.

Tarski may have had in mind some common intuitions of unsoundness, such as the intuitions about sentences of the form 'If *A*, then, if *B* then *A*' mentioned above, when he spoke of the discrepancies between the common concept of consequence and earlier technical concepts of consequence—presumably the derivational concept. But he also speaks of "sceptics" (1983b, p. 410, 1936b, p. 2), who, presumably on more theoretical grounds, had the view that the classical derivational concept, perhaps besides being unsound, was incomplete. It seems likely that he was thinking here of intuitionists, who thought not only that the classical derivational concept was unsound, but also that no derivational concept could exhaust the richness of intuitive logical reasoning (which, like the logicists, they identified with mathematical reasoning). However, Tarski mentions the fact that one could counter these skeptics by pointing out that the derivational logicians "had actually succeeded in reproducing in the shape of formalized proofs all the exact reasonings which had ever been carried out in mathematics" (1983b, p. 410, 1936b, p. 2), thus implying that the skeptics had not been able to substantiate their skepticism with concrete counterexamples to the hypothesis that the classical derivational concept grasped the "essential content" of the common concept of correct argument.

Tarski's considerations on ω-arguments lead him to say that "the scepticism was quite justified" (1983b, p. 410, 1936b, p. 2) after all. Why did he think of the ω-argument he had in mind as valid in virtue of the nature of the common concept of consequence? The

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

probable answer is: because it possesses all the basic intuitive properties we just mentioned two and three paragraphs back. Recall the argument $A'_0, A'_1, \ldots, A'_n, \ldots / A'$ from section 3. First, this is certainly a truth-preserving argument. Second, the argument is an instance of the form

$$\begin{array}{l} P(0), \\ P(1), \\ \ldots \\ P(n), \\ \ldots \\ \hline \forall X^3_1(\text{NATURAL NUMBER}(X^3_1) \supset P(X^3_1)), \end{array}$$

shared by all the arguments in which the part of the original argument $A'_0, A'_1, \ldots, A'_n, \ldots / A'$ that is not directly responsible for its correctness (the logical predicate '$\exists X^1_1 \sim X^2_1(X^1_1)$') is replaced with a grammatically congruous part; and all these arguments will intuitively be truth-preserving too. Third, there is intuitively no unnamed property that can provide a counterexample when thought of as the extension of 'P'. Of course, in order to reach these conclusions we must have the skills necessary to understand the ω-argument Tarski had in mind, so these conclusions could not be elicited from an untrained or an unreflective user of the common concept of consequence. But the conclusions have as their ultimate source the intuitions that form part of the common concept, and are derived from them.

The material inadequacy of the classical derivational concept was the main fact that led Tarski to propose his defined concept of logical consequence. However, there were other motivations that formed part of Tarski's pretheoretic ideas about consequence, and that played a decisive role in shaping the form his defined concept took. Tarski acknowledges that his

> proposed treatment of the concept of consequence makes no very high claim to complete originality. The ideas involved in this treatment will certainly seem to be something well known, or even something of his own, to many a logician who has given close attention to the concept of consequence and has tried to characterize it more precisely. (Tarski, 1983b, p. 114, 1936b, p. 6)

However, he continues, "it seems to me that only the methods which have been developed in recent years for the establishment of scientific semantics, and the concepts defined with their aid, allow us to present these ideas in an exact form" Tarski 1938b, p. 114, 1936b, p. 6. Tarski is implying that a comparatively precise understanding of the notion of consequence has been in the minds of many logicians, presumably the logicians who had used in their researches informal, often primitive, notions of interpretation, model, and truth in a range of interpretations. (In 1935a, pp. 363–364, n. 83, in a note to a passage reminiscent of the one just quoted, which we already mentioned in subsection 2.2, Tarski mentions among these logicians Ackermann (1928), Bernays and Schönfinkel (1928), Herbrand (1930), Löwenheim (1915), and Skolem (1919, 1920, 1929), and he has just mentioned, at p. 102, n. 81, Gödel, 1930.) Many logicians, and especially those intellectually closer to the axiomatic tradition, had used informal but solid ideas about the concepts of interpretation,

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

model of an axiom system, and validity. However, these ideas were "inexact" (by Tarski's high standards); the concepts used had been taken as primitive, or had been ultimately defined in terms of nonmathematical primitives. One of Tarski's motivations was to accommodate this comparatively inexact practice, constructing surrogates of those informal concepts definable within classical mathematics, surrogates with the help of which the results that had been proved about the informal concepts could be reconstructed in an unobjectionably mathematical way.

This was a motivation not dictated by intuitions concerning the content of the common concept of consequence, but not for this reason a less important part of Tarski's pretheoretic conception of what a precise definition of logical consequence should accomplish. There were other related motivations. Tarski was clearly interested not just in grasping the content of the common concept of consequence, but also in doing so in a way that fitted well with as many as possible of the several technical uses for the concept. These uses spanned different traditions, notably the axiomatic and the logicist, but also the algebraic (from which Löwenheim came); and while each had its own idiosyncrasies, a definition of an appropriate technical concept of consequence ought ideally to be of use for all of them, if necessary with the help of conventions suitable to each case.

There was one especially notorious difference between the logicist tradition on the one hand and both the algebraic and the axiomatic traditions on the other. While the logicists saw the (first-order) quantifiers as pure constant terms that always ranged over the same domain (often taken to be the set of all individuals in the world), both the algebraic and the axiomatic schools comfortably worked under the idea that the range of quantification could vary from one interpretation of a mathematical language to another. The logicist view of the quantifiers is probably one of the roots for the frequent logicist treatment of "cardinality assumptions" about the set of all individuals in the world (the set of all "logical individuals") as questions that could or ought to be settled by logic. The axiomatic view of the quantifiers certainly underlies the tacit view that "cardinality assumptions" about the set of specific individuals that one's theory is about could not and ought not to be settled by logic. The right "cardinality assumptions" vary depending on what objects we are concerned with, so for an axiomatic logician they should not be settled once and for all by logic. On the other hand, for a logicist "cardinality assumptions" about "logical" individuals are a stable feature of thought, which is always concerned with the same objects.

It seems clear to me that Tarski sought to make his definition compatible with these different idiosyncrasies. He undoubtedly sought to make his defined concept of logical consequence of use for the axiomatic tradition, as one can gather especially vividly from his frequent references to the tradition in his 1937 manual, together with his presentation of the informal ideas on interpretations, models, and consequence using virtually the same words as in his 1936 consequence paper. In particular, he sought to accommodate the axiomatic tradition's use of different specific domains and its attendant rejection of logically settled "cardinality assumptions". He did this by defining an interpretation as a sequence of extensions for the specific or nonlogical terms of a language or axiomatic theory, but requiring additionally that the domain of quantification be named by the same nonlogical predicate in all interpretations. As Tarski amply illustrates in his manual, the definition and the additional convention can be used to reconstruct the main informal ideas and results of the axiomatic tradition.

But there can also be little doubt that Tarski sought to make his definition of consequence compatible with the idiosyncrasy of the logicist tradition. He makes an obvious effort to show that his defined concept of *analyticity* (or truth in all interpretations) is

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

equivalent with Carnap's (1934, 1937) notion of analyticity, particularly in the special case of Carnap's logicist "language II" (cf. the remarks in TLC, 145f., and, for a detailed account, De Rouilhan, forthcoming); and his example of a particular ω-argument, showing the non-coextensionality between the derivational concept of consequence and the core of the common concept, belongs to a logicist language outside the scope of the axiomatic tradition. The logicist tradition's use of a unique domain for logicist languages and its attendant embrace of logically settled "cardinality assumptions" is again accommodated by Tarski's definition, this time without special conventions. Since a typical logicist language does not contain nonlogical constants, Tarski's definition of an interpretation entails that a language of this sort has only one interpretation. (In the case of the hybrid languages in which some nonlogical constants are superimposed on a purely logical logicist language, but no nonlogical predicate names the domain of "logical" individuals, the definition of an interpretation still entails that all interpretations share the same domain of quantification.) The end result is the desired one that in these cases the "cardinality assumptions" true in the intended domain are declared logical truths, or logical consequences of every set of sentences, by Tarski's definition.

Tarski's pretheoretic conception of consequence had many notes, many sides. It was grounded on fundamental intuitions about the ordinary concept of consequence preexisting in everyday use, but it also incorporated a number of more theoretic suitability requirements inspired by a variety of traditions and contexts of investigation. It is a tribute to Tarski's genius that it is becoming increasingly clear how he managed to accommodate highly successfully all these many ideas in the definition of his model-theoretic concept of logical consequence.

§6. Acknowledgments. Parts of this paper were presented as talks at the UNAM, the Universidade Federal do Rio de Janeiro, the Universidad Nacional Mayor de San Marcos in Lima, the University of California at Berkeley, the Institut d'Histoire et Philosophie des Sciences et des Techniques in Paris, and at the XXII World Congress of Philosophy in Seoul. My warm gratitude to all these audiences for their comments. I also wish to thank especially Bill Hanson, Paolo Mancosu and Jan Woleński for their help.

BIBLIOGRAPHY

Ackermann, W. (1928). Über die Erfüllbarkeit gewisser Zählausdrücke. *Mathematische Annalen*, **100**, 638–649.

Bays, T. (2001). On Tarski on models. *Journal of Symbolic Logic*, **66**, 1701–1726.

Benacerraf, P., & Putnam, H. (editors.). (1983). *The Philosophy of Mathematics* (second edition). Cambridge: Cambridge University Press.

Bernays, P., & Schönfinkel, M. (1928). Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen*, **99**, 342–372.

Carnap, R. (1934). *Logische Syntax der Sprache*. Vienna: Julius Springer.

Carnap, R. (1937). *The Logical Syntax of Language*. London: Routledge & Kegan Paul.

Carnap, R. (1942). *Introduction to Semantics*. Cambridge, MA: Harvard University Press.

Church, A. (1956). *Introduction to Mathematical Logic I*. Princeton, NJ: Princeton University Press.

De Rouilhan, P. (forthcoming). Carnap on logical consequence for languages I and II. In Wagner, P., editor. *Carnap's Logical Syntax of Language*. New York: Palgrave-Macmillan.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

Editorial Office of *Fundamenta Mathematicae*. (1934). Bemerkung der Redaktion. *Fundamenta Mathematicae*, **23**, 161.

Edwards, J. (2003). Reduction and Tarski's definition of logical consequence. *Notre Dame Journal of Formal Logic*, **44**, 49–62.

Etchemendy, J. (1988). Tarski on truth and logical consequence. *Journal of Symbolic Logic*, **53**, 51–79.

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

García-Carpintero, M. (1993). The grounds for the model-theoretic account of the logical properties. *Notre Dame Journal of Formal Logic*, **34**, 107–131.

García-Carpintero, M. (2003). Gómez-Torrente on modality and Tarskian logical consequence. *Theoria* (San Sebastián), **18**, 159–170.

Gödel, K. (1930). Die Vollständigkeit der Axiome der logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, **37**, 349–360. Reprinted in Gödel (1986, pp. 102–122). Translated as "The completeness of the axioms of the functional calculus of logic" in van Heijenoort (1967, pp. 583–591) and in Gödel (1986, pp. 103–123).

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, **38**, 173–198. Reprinted in Gödel (1986, pp. 144–194). Translated as "On formally undecidable propositions of *Principia Mathematica* and related systems I" in van Heijenoort (1967, pp. 596–616) and in Gödel (1986, pp. 145–195).

Gödel, K. (1986). *Collected Works*, Vol. I. New York: Oxford University Press.

Gómez-Torrente, M. (1996a). Tarski on logical consequence. *Notre Dame Journal of Formal Logic*, **37**, 125–151.

Gómez-Torrente, M. (1996b). *Tarski's Definition of Logical Consequence. Historical and Philosophical Aspects* (PhD dissertation, Princeton University). Ann Arbor, MI: University Microfilms.

Gómez-Torrente, M. (1998). On a fallacy attributed to Tarski. *History and Philosophy of Logic*, **19**, 227–234.

Gómez-Torrente, M. (1998/1999). Logical truth and Tarskian logical truth. *Synthèse*, **117**, 375–408.

Gómez-Torrente, M. (2000). A note on formality and logical consequence. *Journal of Philosophical Logic*, **29**, 529–539.

Gómez-Torrente, M. (2004). The indefinability of truth in the *Wahrheitsbegriff*. *Annals of Pure and Applied Logic*, **126**, 27–37.

Hanson, W. H. (1997). The concept of logical consequence. *Philosophical Review*, **106**, 365–409.

Hempel, C. G. (1945). On the nature of mathematical truth. *The American Mathematical Monthly*, **52**, 543–556. Reprinted in Benacerraf and Putnam (1983, pp. 377–393).

Herbrand, J. (1930). Recherches sur la théorie de la démonstration. *Travaux de la Société des Sciences et des Lettres de Varsovie*, Classe III. Sciences mathématiques et physiques, no. 33. Chapter 5 translated as "Investigations in proof theory: The properties of true propositions" in van Heijenoort (1967, pp. 525–581).

Hilbert, D., & Ackermann, W. (1928). *Grundzüge der theoretischen Logik*. Berlin: Julius Springer.

Jané, I. (2006). What is Tarski's *common* concept of consequence?. *Bulletin of Symbolic Logic*, **12**, 1–42.

Lewis, C. I. (1918). *A Survey of Symbolic Logic*. Berkeley, CA: University of California Press.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

Lewis, C. I., & Langford, C. H. (1932). *Symbolic Logic*. New York: Century Company.
Löwenheim, L. (1915). Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, **76**, 447–470. Translated as "On possibilities in the calculus of relatives" in van Heijenoort (1967, pp. 228–251).
Mancosu, P. (2006). Tarski on models and logical consequence. In Ferreirós, J., and Gray, J. J., editors. *The Architecture of Modern Mathematics*. Oxford: Oxford University Press, pp. 209–237.
Prawitz, D. (1985). Remarks on some approaches to the concept of logical consequence. *Synthèse*, **62**, 153–171.
Ramsey, F. P. (1925). The foundations of mathematics. *Proceedings of the London Mathematical Society*, 2nd series, **25**, 338–384. Reprinted in Ramsey (1931, pp. 1–61).
Ramsey, F. P. (1931). *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul.
Ray, G. (1996). Logical consequence: A defense of Tarski. *Journal of Philosophical Logic*, **25**, 617–677.
Shapiro, S. (1998). Logical consequence: Models and modality. In Schirn, M., editor. *The Philosophy of Mathematics Today*. Oxford: Oxford University Press, pp. 131–156.
Sher, G. (1991). *The Bounds of Logic*. Cambridge, MA: M.I.T. Press.
Sher, G. (1996). Did Tarski commit 'Tarski's fallacy'?. *Journal of Symbolic Logic*, **61**, 653–686.
Skolem, T. (1919). Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassem von Aussagen betreffen. *Videnskapsselskapets Skrifter*, I Mat.-nat. Klasse, no. 3.
Skolem, T. (1920). Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Videnskapsselskapets Skrifter*, I Mat.-nat. Klasse, no. 4. Section 1 translated as "Logico-combinatorial investigations in the satisfiability or provability of mathematical propositions: A simplified proof of a theorem by L. Löwenheim and generalizations of the theorem" in van Heijenoort (1967, pp. 252–263).
Skolem, T. (1929). Über einige Grundlagenfragen der Mathematik. *Skrifter utgitt av Det Norske Videnskaps-Akademi I Oslo*, I Mat.-nat. Klasse, no. 4.
Soames, S. (1999). *Understanding Truth*. New York: Oxford University Press.
Stroińska, M., & Hitchcock, D. (2002). Introduction to Tarski (2002, pp. 155–175).
Tarski, A. (1931). Sur les ensembles définissables de nombres réels. I. *Fundamenta Mathematicae*, **17**, 210–239.
Tarski, A. (1933). Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik*, **40**, 97–112.
Tarski, A. (1935a). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, **1**, 261–405.
Tarski, A. (1935b). Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis*, **5**, 80–100.
Tarski, A. (1936a). O pojęciu wynikania logicznego. *Przegląd Filozoficzny*, **39**, 58–68. References to the English translation Tarski (2002).
Tarski, A. (1936b). Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique*, fasc. 7 (Actualités Scientifiques et Industrielles, Vol. 394). Paris: Hermann et Cie, pp. 1–11.
Tarski, A. (1936c). *O logice matematycznej i metodzie dedukcyjnej*. Lvov-Warsaw: Książnica-Atlas.

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press

Tarski, A. (1936d). Grundzüge des Systemenkalküls, Zweiter Teil. *Fundamenta Mathematicae*, **26**, 283–301.
Tarski, A. (1937a). *Einführung in die mathematische Logik und in die Methodologie der Mathematik*. Vienna: Springer.
Tarski, A. (1937b). Sur la méthode déductive. In *Travaux du IX^e Congrès International de Philosophie*, tome 6 (Actualités Scientifiques et Industrielles, Vol. 535). Paris, France: Hermann et Cie, pp. 95–103.
Tarski, A. (1940). On the completeness and categoricity of deductive systems. Unpublished typescript, Alfred Tarski Papers, Carton 15, Bancroft Library, U.C. Berkeley. A version of the typescript edited by Paolo Mancosu is to appear as an archival appendix to Mancosu's book *The Adventure of Reason*, forthcoming from Oxford University Press.
Tarski, A. (1941). *Introduction to Logic and to the Methodology of Deductive Sciences*. New York: Oxford University Press.
Tarski, A. (1946). *Introduction to Logic and to the Methodology of Deductive Sciences* (second edition). New York: Oxford University Press.
Tarski, A. (1953). "A general method in proofs of undecidability" in Tarski *et al.* (1953, pp. 1–35).
Tarski, A. (1965). *Introduction to Logic and to the Methodology of Deductive Sciences* (third edition). New York: Oxford University Press.
Tarski, A. (1983a). *Logic, Semantics, Metamathematics* (second edition, J. Corcoran, editor). Indianapolis, IN: Hackett.
Tarski, A. (1983b). On the concept of logical consequence. Translation of Tarski (1936b) by Woodger, J. H. in Tarski (1983a, pp. 409–420).
Tarski, A. (2002). On the concept of following logically. Translation of Tarski (1936a) by Stroińska, M., & Hitchcock, D. *History and Philosophy of Logic*, **23**, 155–196.
Tarski, A., & Lindenbaum, A. (1935). Über die Beschränktheit der Ausdrucksmittel deduktiver Theorien. *Ergebnisse eines mathematischen Kolloquiums*, fasc. 7, 1934–1935, 15–22.
Tarski, A., Mostowski, A., & Robinson, R. (1953). *Undecidable Theories*. Amsterdam, The Netherlands: North-Holland.
van Heijenoort, J. (editor). (1967). *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931*. Cambridge, MA: Harvard University Press.
Whitehead, A. N., & Russell, B. (1925–1927). *Principia Mathematica* (second edition, 3 vols). Cambridge: Cambridge University Press.

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO DF 04510, MEXICO
E-mail: mariogt@servidor.unam.mx

https://doi.org/10.1017/S1755020309090200 Published online by Cambridge University Press