J Log Lang Inf (2010) 19:315–326
DOI 10.1007/s10849-009-9103-z

Circularity or Lacunae in Tarski's Truth-Schemata

Dale Jacquette

Published online: 1 November 2009

Abstract Tarski avoids the liar paradox by relativizing truth and falsehood to particular languages and forbidding the predication to sentences in a language of truth or falsehood by any sentences belonging to the same language. The Tarski truth-schemata stratify an object-language and indefinitely ascending hierarchy of meta-languages in which the truth or falsehood of sentences in a language can only be asserted or denied in a higher-order meta-language. However, Tarski's statement of the truth-schemata themselves involve general truth functions, and in particular the biconditional, defined in terms of truth conditions involving truth values standardly displayed in a truth table. Consistently with his semantic program, all such truth values should also be relativized to particular languages for Tarski. The objection thus points toward the more interesting problem of Tarski's concept of the exact status of truth predications in a general logic of sentential connectives. Tarski's three-part solution to the circularity objection which he anticipates is discussed and refuted in detail.

Keywords Circularity · Language, meta-language · Semantics · Tarski, Alfred · Truth

1 Tarski's Semantic Conception of Truth

To avoid the liar paradox, Alfred Tarski proposes to relativize truth predications to particular formalized languages. Assertions and denials of the truth of sentences are restricted from being made within the same language to which the sentences belong, but are permitted only in higher-order meta-languages (Tarski 1983).

Tarski's truth-schemata are not intended to provide an analysis of the general concept of truth, but rather of truth-in-a-formalized-language-L. The schemata can be

D. Jacquette (✉)
University of Bern, Bern, Switzerland
e-mail: dale.jacquette@philo.unibe.ch

formulated in this fashion, organized in an indefinitely ascending hierarchy of object- and meta-languages, proceeding from the lowest object-language level to increasingly higher meta-languages:

etc.		etc.
⋮		⋮
"'p' is true$_{L1}$ ↔ p' is true$_{L2}$ ↔ ['p' is true$_{L1}$ ↔ p]' is true$_{L3}$ ↔ ["p' is true$_{L1}$ ↔ p' is true$_{L2}$ ↔ ['p' is true$_{L1}$ ↔ p]]	Meta-language	3
"p' is true$_{L1}$ ↔ p' is true$_{L2}$ ↔ ['p' is true$_{L1}$ ↔ p]	Meta-language	2
'p' is true$_{L1}$ ↔ p	Meta-language	1
p	Object-language	O

At the object-language level there is as yet no truth predication, according to Tarski, but only sentences in which no truth or falsehood predicates occur. If the language contained its own truth predicate, contrary to Tarski's restriction, then it would be possible within that language to construct a liar sentence denying its own truth or asserting its own falsehood, and thereby jeopardizing the language's classical bivalent semantic integrity. The simplest form of the Tarski truth schema appears at the level of meta-language 1, in which it is possible to predicate truth or falsehood of sentences in the lower object-language. The Tarski schema, for obvious reasons, is also sometimes misleadingly described as a disquotational, redundancy, or correspondence analysis of truth relative to the meta-language in which truth predications are asserted or denied of sentences in the object-language or lower-order meta-languages in the hierarachy. The identical schema is replicated in all higher-order meta-languages $m + 1$, by uniformly substituting for sentence p in the meta-language 1 truth schema the truth predications permitted in lower-order meta-language m $(m > 1)$.

2 The Circularity Problem

The essentials of Tarskian truth schemata are well-known.[1] A circularity threatens Tarski's proposal for understanding the concept of truth in formalized languages nevertheless, of which Tarski is well aware and which he tries to address.[2]

[1] Here I follow standard practice in symbolizing Tarski's biconditional truth schema, suggested also by Donald Davidson's modification of the principle as the so-called T-schema, 'Truth and Meaning' [1967], reprinted in *Inquiries into Truth and Interpretation* (Oxford: The Clarendon Press, 1984), where Davidson writes, p. 23: 'The theory will have done its work if it provides, for every sentence s in the language under study, a matching sentence (to replace "p") that, in some way yet to be made clear, "gives the meaning" of s. One obvious candidate for matching sentence is just s itself, if the object-language is contained in the meta-language; otherwise a translation of s in the meta-language. As a final bold step, let us try treating the position occupied by "p" extensionally: to implement this, sweep away the obscure "means that", provide the sentence that replaces "p" with a proper sentential connective, and supply the description that replaces "s" with its own predicate. The plausible result is / (T)s is T [true] if and only if p. / What we require of a theory of meaning for a language L is that without appeal to any (further) semantical notions it place enough restrictions on the [truth] predicate "is T" to entail all sentences got from schema T when "s" is replaced by a structural description of a sentence of L and "p" by that sentence.'

[2] Criticisms of Tarski's truth convention are offered by Field (1972), Gupta (1993), Halbach (1999), Hintikka (1975), Ketland (1999).

The problem is that the truth functional biconditional appears innocently in these formalizations as though it were a logically neutral way of relating the truth conditions of propositions relativized to object- and meta-languages in the Tarskian hierarchy. The biconditional itself is nevertheless a relation that is defined and enters into philosophical semantics for interpretation as a truth-function, the very formal interpretation of which presupposes the concept of truth. We see this unmistakably in the ordinary truth table definition by cases of the biconditional, by which a biconditional is true (T) (the biconditional truth function yields output T) just when its component (input) sentences both have truth value true (T) or both have truth value false (F).

It is standard to excuse the use of truth functions like the biconditional in analyzing the concept of truth as meta-logical or meta-linguistic, as belonging more particularly to a meta-language in which the requirements of true propositions are spelled out. Such a ploy is particularly unacceptable, however, in the context of Tarski's explication of truth conditions for propositions in formalized languages, which is already stratified into an object-language and meta-language hierarchy, where it seems to involve Tarski's theory in vicious circularity.

Tarski considers the sentence p ('Snow is white') as belonging to an object-language, and then requires that the sentence in which truth is predicated of p belong to a higher-order meta-language, sanctioning the occurrence of the biconditional ('iff' or $\leftrightarrow$ in formalizations) as common to any truth or falsehood predication in the meta-language hierarchy, as well as in non-atomic object-language sentences that happen to be biconditional in logical form. As we have seen, there is no such thing as simple univocal truth if Tarski's theory is correct, but only truth-in-a-given-meta-language-L_i for a certain object-language or lower-level meta-language L_{i-1}. If there is no such thing as simple univocal truth, nor a general analysis of the concept of truth, however, then there are equally no simple univocal truth functions, for each truth function is defined by reference to truth condition cases, expressed as T and F combinations, as in the standard truth tables for the five most common truth functions— negation, conjunction, inclusive disjunction, material conditional, and biconditional. If Tarski's relativized theory of truth in formalized languages is correct, then it is mistaken to read T in a standard truth table as representing truth simpliciter. We cannot refer to truth, falsehood, or truth values generally, according to Tarski's concept of truth in formalized languages, so we should not be able to do so even in an elementary truth table, if a general analysis of the concept of truth on pain of logical-semantic paradox is supposed to be unavailable. Instead, we can only consistently speak of truth-or-falsehood-in-a-given-meta-language-L_i for the sentences appearing in an object-language or lower-level meta-language L_{i-1}.

Thus, in 'The Concept of Truth in Formalized Languages', Tarski explicitly avails himself of stock truth functional connectives in symbolic logic and their natural language equivalents in order to formulate his truth-schemata. There, for example, he maintains:

> Among the expressions of the metalanguage we can distinguish two kinds. To the first belong *expressions of a general logical character*, drawn from any sufficiently developed system of mathematical logic [Tarski's italics; he refers in the footnote here to Whitehead and Russell's *Principia Mathematica*]…To the

same category belongs a series of analogous expressions from the domain of the sentential calculus, of the first order functional calculus and of the calculus of classes, for example, '*if* …, *then*', '*and*', '*if and only if*', '*for some x*' (or '*there is an x such that* …'), ….[3]

Applying the standard apparatus of sentential or propositional logic, Tarski proceeds to formulate his treatment of truth conditions relative to a language, first intuitively:

In order to explain the sense of this phrase we consider the following scheme:

for all a, a satisfies the sentential function x if and only if p

and substitute in this scheme for 'p' the given sentential function (after first replacing the free variable occurring in it by 'a') and for 'x' some individual name of this function. Within colloquial language we can in this way obtain, for example, the following formulation:

for all a, a satisfies the sentential function 'x is white' if and only if a is white

(and from this conclude, in particular, that snow satisfies the function 'x is white).[4]

And then in terms of a concept of a satisfaction set somewhat more rigorously defined in terms of the calculus of classes:

Definition 25 x is a correct (true) sentence in the individual domain a if and only if $x \in S$ and every infinite sequence of sub-classes of the class a satisfies the sentence x in the individual domain a.[5]

Although the implication has not been widely acknowledged, if strictly carried into practice, Tarski's theory of truth requires a parallel hierarchy of truth table definitions for an indefinitely ascending hierarchy of truth functions. This means that for Tarski there is equally no simple univocal negation, conjunction, disjunction, conditional, or biconditional, but rather, only, for example, $\leftrightarrow$-in-a-given-object-or-meta-language-L_i, for all the object- and meta-languages required by Tarski's concept of truth, $L_o, \ldots, L_n, \ldots$. When Tarski appeals to the biconditional to express the correspondence relation whereby a sentence 'p' is true$_L$ iff $(\leftrightarrow)p$, the relation should be formulated: 'p' is true$_{Li} \leftrightarrow_j p$, where it remains to be seen whether $i = j$ or $i \neq j$ (and, if the latter, whether $i > j$ or $i < j$). Thus, Tarski's definition of the concept of truth-in-a-given-language-L in the above formulation presupposes the very same concept of truth-in-a-given-language-L that is presupposed in the truth table definition of the truth function $\leftrightarrow_L$. As an epistemic indication of the account's circularity, we must already know what it means to speak of true-in-L in order to understand the appropriately linguistically relativized truth function $\leftrightarrow_L$, by which the concept of true-in-L in strict adherence to Tarski's conclusions needs to be defined.

[3] Tarski, The Concept of Truth in Formalized Languages, *Logic, Semantics, Metamathematics*, 170-171.

[4] Ibid, p. 190.

[5] Ibid, p. 200.

3 Tarski's Three-Part Solution

Tarski anticipates precisely this criticism in his 1944 essay, 'The Semantic Conception of Truth and the Foundations of Semantics'. In section II, Polemical Remarks, Tarski describes and attempts to refute what he calls 'a typical example' of an objection 'to the semantic conception of truth in general' (Tarski 1944). Tarski writes:

> In formulating the definition we use necessarily sentential connectives, i.e., expressions like "*if*…, *then*," "or," etc. They occur in the definiens; and one of them, namely, the phrase "*if*, *and only if*" is usually employed to combine the definiendum with the definiens. However, it is well known that the meaning of sentential connectives is explained in logic with the help of the words "*true*" and "*false*"; for instance, we say that an equivalence, i.e., a sentence of the form "*p if*, *and only if*, *q*," is true if either both of its members, i.e., the sentences represented by '*p*' and '*q*,' are true or both are false. Hence the definition involves a vicious circle.[6]

Tarski's reply to the objection has three parts, none of which in the end seems decisive or adequate to withstand higher counter-criticisms.

(1) Tarski begins by trying to allay concern about the objection on the grounds that any effort to formally explicate a concept of truth would find itself in the same sinking boat. He explains: 'If this objection were valid, no formally correct definition of truth would be possible; for we are unable to formulate any compound sentence without using sentential connectives, or other logical terms defined with their help. Fortunately, the situation is not so bad.'[7]

It is a platitude among logicians interested in philosophical argument to say that one thinker's *modus ponens* is another's *modus tollens*. Tarski favors the *modus tollens* side on this issue, although he does not simply conclude that the objection must be invalid on the grounds that we somehow know in advance that a formally correct definition of truth must be possible. Nevertheless, in keeping with the rhetorical burden of the polemical remarks in this section of the essay, Tarski casts a first ray of suspicion on the objection by suggesting that it would be too strong if correct, since it would exclude any proposal for formally defining a concept of truth regardless of the definition's content. The whole enterprise of trying to clarify the semantics of truth in any context and by any means is accordingly placed in doubt.

As already indicated, this is precisely the conclusion disparaging purely formal explications of the concept of truth that we have maintained must be taken seriously, as our *modus ponens* stands in opposition to Tarski's *modus tollens*. The implication then is that there cannot be a formal semantic definition of the concept of truth, a conclusion that can be philosophically supported on a number of grounds. Tarski recognizes the general threat to a formal semantics of truth conditions especially as an adequacy criterion for definitions of truth. The difference is that Tarski believes that he can rescue the project of providing a semantic treatment of the concept of truth

[6] Ibid.

[7] Ibid.

in axiomatized formal languages rather than in a language interpreted in the sense of model theory. To rescue the semantics of truth from the objection that the definition relies on interpreted truth functions Tarski's defense therefore boils down to the remaining two components of his reply.

(2) Tarski next confronts the circularity objection head-on, arguing that definitions of the truth functional connectives in terms of a prior concept of truth is a superfluous meta-linguistic superaddition to a formal logical system considered strictly and correctly as such. He continues:

> It is undoubtedly the case that a strictly deductive development of logic is often preceded by certain statements explaining the conditions under which sentences of the form "*if p*, *then q*," etc., are considered true or false. (Such explanations are often given schematically, by means of the so-called truth-tables.) However, these statements are outside of the system of logic, and should not be regarded as definitions of the terms involved. They are not formulated in the language of the system, but constitute rather special consequences of the definition of truth given in the metalanguage (Tarski 1944).

Tarski's strategy is interesting, but not obviously satisfactory. He acknowledges that truth table definitions of the propositional truth functional connectives often herald a 'strictly deductive development of logic', but he argues that such definitions do not belong to the system of logic whose formal expression they appear to prepare. He invokes the object-language meta-language distinction to preserve a formal system of logic belonging to the object-language from a circularity that he implies can only threaten a semantic concept of truth if object- and meta-languages are thoughtlessly conflated and confused.

That there is a distinction between logic as an axiomatic object-language and its meta-languages is not under dispute. The question is rather whether and in what sense the distinction helps Tarski to avoid the objection that there is a vicious circularity in the use of truth functional propositional connectives to define the concept of truth relativized to formal languages. The objection, first of all, is not that there is a circularity in the object-language of symbolic logic, but that there is a circularity specifically within the meta-language that purports to articulate adequacy criteria for true sentences belonging to the object-language. Tarski declares that truth table definitions of the truth functional propositional connectives are not part of the object-language of logic, and so they are not. However, the circularity of concern in the objection under consideration is internal to the object-language's semantic meta-language in which the concept of truth is supposed to be formally defined for the object-language.

There might be relief from the circularity if we could have a meta-language in which the truth of sentences in an object- or lower-level meta-language were defined that was distinct from the meta-language in which the essential truth functional propositional connective, in particular, the biconditional, $\leftrightarrow$, is defined. However, a moment's reflection shows that the meta-language M_{n+1} of object- or meta-language L_n, in which a truth functional propositional connective is defined for L_n, cannot be distinct from the meta-language in which the truth of sentences belonging to L_n is defined. Thus, if M_{n+1} is the meta-language in which the truth of sentences in object- or meta-language L_n is defined, then the truth functional propositional connectives occurring in L_n must

also be defined in M_{n+1}, and not in some other meta-language $\neq M_{n+1}$. The reason is clear when we appreciate the fact that a propositional connective could not possibly be understood as a truth function involving sentences belonging to another language than that for which the connective is defined.

Suppose, then, as before, that M_{n+1} is the meta-language of object- or meta-language L_n, and that $\leftrightarrow$ is defined for L_n in M_{n+1}. When we ask whether the truth conditions for sentences connected by $\leftrightarrow$ in L_n could be defined in some other language than M_{n+1}, it should be immediately clear that the answer is no; for if they were, then even in principle the potential truth values of sentences connected by $\leftrightarrow$ would be unavailable for input to the truth function within M_{n+1}. If truth values and the definitions of truth functional connectives are relativized to specific formalized languages, as Tarski proposes, then no language can define a connective such as $\leftrightarrow$ in a formalized language L_n if it cannot at the same time make reference to available language-relativized truth conditions for sentences belonging to L_n.

If we try to save the situation by maintaining that M_{n+1} could be distinct from another meta-language M_{n+2} or M^*_{n+1} of L_n, which, as the alternative notation indicates, may or may not be stratified relative to M_{n+1}, but are in any case ostensibly different from it, in which the truth conditions for sentences connected by $\leftrightarrow$ in L_n are separately defined but still available to M_{n+1} for purposes of defining $\leftrightarrow$ and other truth functional propositional connectives in L_n, then M_{n+1} effectively contains all the relevant information vouchsafed by M_{n+2} or M^*_{n+1}, and in that respect M_{n+1} and M_{n+2} or M^*_{n+1} relevantly coincide. There is but one meta-language of L_n in that case and at least in that limited respect through which $\leftrightarrow$ is defined for sentences in L_n. Such an overlap of available information and one-way referencing concerning the semantic status of sentences in a language subordinate to a meta-language in which a propositional connection such as $\leftrightarrow$ is nevertheless sufficient to embroil Tarski in the circularity objection he is hoping to avoid. Tarski believes that the hierarchy of object- and meta-languages provides a route of escape from the circularity objection, but he seems not to consider the exact information required within a meta-language in order to define the relevant semantic properties of a subordinate language, which is all that is needed for the circularity problem to arise.

(3) Tarski next offers another defense of his semantic conception of truth in response to the circularity objection. He now maintains:

> Moreover, these statements do not influence the deductive development of logic in any way. For in such a development we do not discuss the question of whether a given statement is true, we are only interested in the problem whether it is provable.[8]

Tarski, interestingly, also maintains in this third part of his defense against the circularity objection that in the deductive development of logic, as quoted above, 'we do not discuss the question of whether a given sentence is true, we are only interested in the problem whether it is provable'.

[8] Ibid.

There is sense in which Tarski is obviously correct, particularly if we think of an axiomatized deductive system in purely formalist terms entirely as an uninterpreted symbol manipulation game. There is much to recommend such an approach to logic, just as there is in Hilbertian philosophy of mathematics. Requirements of provability can be stipulatively configured for a logic's syntax, with no consideration for the actual truth values of the propositions that enter into deductive derivations of one set of wffs from another.

If 'deduction' is intended here as it generally meant in logical theory, however, then there is another respect in which the concept of truth enters indispensably into the theory and practice of deductively valid inference. We cannot intelligibly define deduction as any rule-governed syntax transformation we please, for then we could allow any recognized inferential fallacy as a deductively valid proof. Instead, the formal derivation of sentences within a formalized language is constrained by the necessity of being truth-preserving. Thereby the concept of truth, including its applications in syntactically considered deductively valid proof, involving the biconditional among other truth functional propositional connectives, defined in effectively the same meta-language or relevantly overlapping fragments of distinct meta-languages referentially or in other ways informationally interconnected, remains ineluctably entangled in vicious circularity.

The fact, if it is a fact, that we appear to be interested only in the question of provability in the deductive development of logic does not thereby remove the concept of truth from consideration of the deductively valid derivation of proofs. It is only by virtue of and with tacit reference to our background understanding of the truth-preserving requirements of deductively valid inference that we can intelligibly speak of proof and provability in formalized languages. If we eliminate the concept of truth, then we equally eliminate the possibility of deductively valid proof.

Tarski has it partly right when he concludes in this part of his countercriticism of the circularity objection:

> …the moment we find ourselves within a deductive system of logic — or of any discipline based upon logic, e.g., of semantics — we either treat sentential connectives as undefined terms, or else we define them by means of other sentential connectives, but never by means of semantic terms like "*true*" or "*false*". For instance, if we agree to regard the expressions "*not*" and "*if…, then*" (and possibly also "*if, and only if*") as undefined terms, we can define the term "*or*" by stating that a sentence of the form "*p or q*" is equivalent to the corresponding sentence of the form "*if not p, then q*." The definition can be formulated, e.g., in the following way:
>
> *(p or q) if, and only if, (if not p, then q)*
>
> This definition obviously contains no semantic terms.[9]

What Tarski says here is perfectly true, but it does not necessarily support the conclusion he needs to sustain in order to avoid the circularity objection. It is correct

[9] Ibid.

to observe that we do not as a rule explicitly use, although neither are we prevented from explicitly using, semantic terminology in working competently with the usual sentential connectives.

Looking anthropologically at what logicians actually do in practice, it is reasonable to remark as Tarski does that sentential connectives are treated either as undefined terms or else defined in relation to other sentential connectives. The fact that we take these purely syntactical shortcuts in using a logical symbolism need not stand in doubt. The deeper question is nevertheless whether we can possibly be justified in doing so without at least implicit reference to the truth conditions of propositions whereby the syntactical equivalences among the propositional connectives are legitimized.

After we become familiar with the use of a logical symbolism we can use its terms and operators without thought as to their semantic underpinning. The propositions formulated in a logical symbolism are nevertheless interrelated by virtue of their meaning. It is surely no accident that the development of logical formalisms generally takes the trouble to introduce the propositional connectives by means of truth tables, to which appeal can be made at any stage as a check on illicit inferences and syntax transformations. If absolutely all of the propositional connectives are treated as undefined, then there is no point at which meaning enters the symbolism, and no point at which we can explain why the material conditional and disjunction with negation can be logically interdefined. If we try to say that the equivalence Tarski mentions, $[p \vee q] \leftrightarrow [\neg p \rightarrow q]$, is purely stipulative, then, setting aside the trivial and universally acknowledged conventionality of one over another choice of notation, we deprive logic of its semantic grounding as a system of functions on true or false propositions.

There is a *reason* why $[p \vee q] \leftrightarrow [\neg p \rightarrow q]$ supports the replacement and transformation of $p \vee q$ by and to $\neg p \rightarrow q$, and not by or to $p \rightarrow q$ or $\neg p \wedge \neg q$, and so on, and the reason has to do with the definition of the propositional connectives in terms of their truth conditions. Whether or not we mention these foundational semantic relations in getting on with the business of using formal symbolic logic to develop a deductive system and its applications, after we have developed a level of comfort and facility with the formalism, does not change the fact that propositional logic contains no equivalences, contrary to Tarski's formalistic avowal, if either all of the propositional connectives are not merely treated as but are in fact undefined terms or reductively related in a chain of transformational syntactical equivalences that ultimately rest on undefined terms. Logic at its propositional root is about the interrelation of possibilities among the truth-valued expressions of thought, and as such is implicitly throughout its superstructure permeated by the concept of truth.

Finally, Tarski offers the following general considerations on the requirements for a definition to embody a vicious circularity:

> However, a vicious circle in definition arises only when the definiens contains either the term to be defined itself, or other terms defined with its help. Thus we clearly see that the use of sentential connectives in defining the semantic term "*true*" does not involve any circle.[10]

[10] Ibid.

This, too, is absolutely true as far as it goes. The trouble is that when we leave truth out of the picture entirely, not merely relying on our informal background understanding of the truth conditions of the propositional connectives in considering the meaning of efforts to explicate the semantics of truth, then we are left at best with an incomplete account of the concept's meaning. Lacking this, we can reasonably say that the concept has not actually been defined at all. If relying exclusively on formal methods in logic we find ourselves unable to define a concept such as the truth of a sentence in a formalized language, then we should be mindful of the limitations and skeptical about the prospects of elaborating a purely formal semantic concept of truth.

Consider the analogous situation in which we try to define 'A' in terms of 'B' in $A =_{df} B$, but where 'B' itself is altogether meaningless. We may try to take comfort, as Tarski evidently does, in the fact that the *definiens* in a formal semantic definition of truth relativized to a formalized language does not contain the term featured in the *definiendum*, or 'other terms defined with its help'. If, however, we cannot fully or properly understand the *definiens* without appealing at least implicitly to an understanding of the concept represented by the *definiendum* term, then we are equally ensnared in vicious circularity.

4 Limitations of Purely Formal Truth Definitions

What are the implications of Tarski's use of truth functions to define a semantic conception of truth? The circularity that appears to threaten Tarski's project has not seemed to worry subsequent generations of logicians and philosophers of language. The explanation may have to do in part with a widespread faith that a solution must be available in a proper application of the object-language and meta-language distinction, and in the fact that Tarski was aware of the difficulty and had answered it satsifactorily. Yet the problem is not easily dismissed and poses a genuine threat to any effort to provide a purely formal definition of truth.

Struggling against the manifest vacuity of a mere coherence net of inter-defined terms, in which a vicious circularity is rightly sensed, Tarski resists, not, like other theorists, by invoking a category of unexplicated primitive concepts, whose meaning is informally understood, and whose application might be conveyed in other ways outside the system of definitions. Such a solution is unacceptable to Tarski, and methodologically unavailable to his philosophical project of providing a purely formal semantic characterization of truth in formalized languages, which cannot rely on what we know informally about basic concepts behind the scenes. Instead, he proposes to avoid the specter of circularity by boldly appealing to concepts that he maintains have no meaning at all within the specific language in which they occur and in which their definition appears. If this is well in keeping with Tarski's formalist ideology, then, we might conclude, so much the worse for formalism. Tarski may have consistency on his side, but at its altar he sacrifices completeness. The truth-valuational meaning of the propositional connectives is supposed to be only meta-linguistically definable, but we have seen that invoking the object- versus meta-language ploy offers no safe escape from vicious implicit circularity in Tarski's use of truth functions in the semantic definition of truth. It is hard to appreciate how the general expectation that *ex nihilo nihil*

fit is supposed to find exception in the effort to define a semantic conception of truth as relying on terms that are literally meaningless within the language in which the definition is formulated.

Tarski cannot hope to avoid commitment to a semantic account of truth functions, for they are, after all *truth* functions, and Tarski recognizes the need to articulate a semantic conception of truth. He proposes adequacy criteria for a definition of truth as preparation for his definition of truth as a semantic concept. Truth he defines by means of the biconditional, and although the word 'true' does not explicitly appear in the *definiens*, as Tarski rightly remarks, the concept of truth is presupposed by the proper application of the biconditional as a truth function, a function with truth valued propositions as inputs and outputs. We cannot break the circle by trying to treat the propositional connectives as uninterpreted purely syntactical inscription types, nor does the object- versus meta-language distinction avail, for the circularity belongs entirely to the meta-language. The circularity is inherent in any effort to work out a purely formal semantic definition of truth, and the purely syntactical theory of truth functions is perhaps the least convincing component of Tarski's program. It demonstrates a desperate but ultimately unsuccessful attempt to cope with the circularity involved in using truth functions or truth value defined propositional connectives to define the concept of truth. The same problem does not similarly affect other non-purely-formal efforts to define the concept of truth, as a contribution to the larger project of developing a semantics for formal symbolic logic. It is rather an encumbrance specific to the kind of definition of truth that Tarski proposes, in which adequacy conditions are spelled out exclusively formally in terms of the correspondence between names for sentences and the facts the sentences represent.

Nor can Tarski escape from the problem by stipulating that 'p' is true$_{Li} \leftrightarrow_{Lj} p$, where $i \neq j$. The reason is that, intuitively, if sentence p belongs to language L_i, then, on the assumption that there are no universally general truth functions just as there is no universally general concept of truth, any sentence logically equivalent to p must also belong to L_i. However, p is logically equivalent, as we would express it outside of a Tarskian truth hierarchy framework, to $p \leftrightarrow p$, and this in turn is logically equivalent to $(p \leftrightarrow p) \leftrightarrow p$, and so on, indefinitely. In Tarski's semantic hierarchy, making all the appropriate linguistic and meta-linguistic relativizations explicit, we would need to write out these presumed truth-functional logical equivalences as: $p \leftrightarrow_{Li} p$, $(p \leftrightarrow_{Li} p) \leftrightarrow_{Lj} p$, $((p \leftrightarrow_{Li} p) \leftrightarrow_{Lj} p) \leftrightarrow_{Lk} p$, etc., indefinitely.

The implication is that for Tarski there can be no single univocal higher-level meta-language in which to attribute truth to all of these equivalences. In a sense, therefore, we cannot reach high enough to capture a meta-language that serves all of these equivalences to p in a single univocal truth predication, since each biconditional takes us another step higher, even though p itself belongs to a lower-level formal language, and possibly to the object-language. Tarski nevertheless presents the truth-schemata using truth functions defined by means of truth conditions involving truth values with blithe unconcern, as though in the cases of sentential connectives there was no need to worry about the relativization of truth that he otherwise insists must be observed in order to avoid logical antinomy. Tarski could circumvent the limitation only by imposing a univocal biconditional truth function that does not need to be relativized to any

specific object- or meta-language, but stands outside of all of them simply by virtue of being a part of a general logic. Such a stipulation on Tarski's part, unfortunately, would not only be inconsistent with Tarski's hierarchy of truth-predication languages, but would introduce precisely the vicious circularity in the use of truth-defined truth functions like the biconditional to explicate the concept of truth in a formalized language such as symbolic logic.[11]

References

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: The Clarendon Press.
Field, H. (1972). Tarski's theory of truth. *The Journal of Philosophy, 69*, 347–375.
Gupta, A. (1993). A critique of deflationism. *Philosophical Topics, 21*, 57–81.
Halbach, V. (1999). Disquotationalism and infinite conjunctions. *Mind, 108*, 1–22.
Hintikka, J. (1975). A counterexample to Tarski-type truth-definition as applied to natural languages. *Philosophia, 5*, 207–212.
Ketland, J. (1999). Deflationism and Tarski's paradise. *Mind, 108*, 69–94.
Tarski, A. (1923). On the primitive term of logistic. *Logic, Semantics, Metamathematics*, pp 1–23.
Tarski, A (1935a). Foundations of the calculus of systems. *Logic, Semantics, Metamathematics*, pp. 342–383.
Tarski, A. (1935b). On extensions of incomplete systems of the sentential calculus. *Logic, Semantics, Metamathematics*, pp. 393–400.
Tarski, A. (1937). Sentential calculus and topology. *Logic, Semantics, Metamathematics*, pp. 421–454.
Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research, 4*, 341–375.
Tarski, A. (1983). The concept of truth in formalized languages. In Tarski A (Ed.), *Logic, semantics, metamathematics: Papers from 1923 to 1938 (J. H. Woodger, trans)*, 2nd edn, edited by J. Corcoran, (pp. 152–278). Indianapolis: Hackett Publishing Company.

[11] Tarski (1923) evinces no recognition of the fact, even independently of the circularity objection, that the elementary truth functions, defined in terms of the truth values T and F need to be relativized to particular formalized languages as much as he maintains they do in truth predications attaching to sentences in applications outside of truth table definitions. It is unsurprising that he does not do so in his 1923 essay, 'On the Primitive Term of Logistic', *Logic, Semantics, Metamathematics*, pp. 1–23, based on his dissertation thesis of the same year. However, it is more curious and arguably less excusable to see that he does not do so in writings appearing after his 1930 'The Concept of Truth in Formalized Languages', reporting results Tarski states he achieved in 1929, such as his 1935 essays, 'Foundations of the Calculus of Systems' and 'On Extensions of Incomplete Systems of the Sentential Calculus', *Logic, Semantics, Metamathematics*, pp. 342–383; 393–400 (using unrelativized Polish notation for standard sentential connectives), and 1937 essay, 'Sentential Calculus and Topology', *Logic, Semantics, Metamathematics*, pp. 421–454.