The Semantic Conception of Truth

Anders Johan Schoubye

Bachelor's Paper
Undergraduate Studies in Philosophy, University of Copenhagen
Dept. Philosophy, Education, and Rhetoric
e-mail: Schoubye@hum.ku.dk

June, 2003

An exposition of Tarski's 'Semantic Conception of Truth'
Including a subsequent discussion of various objections put forward

INTRODUCTION

This paper intends to elucidate the definition of truth put forward by Alfred Tarski in 1933. I have deliberately avoided an opening section on truth theories in general, as it would turn out superficial and rather unimportant in view of understanding Tarski's semantic conception of truth. Therefore, no attention is given to theories such as the correspondence theory, the coherence theory, Peirce's pragmaticism, Ramsey's redundancy theory and so on.

Instead I have intended to provide a rather broad explanation of Tarski's semantic definition of truth. However, giving a thorough and concise explanation of Tarski's renowned truth definition has been a complicated task. I was somewhat surprised with the subject's level of difficulty and in addition the amount of literature covering the issue.

I found the inclusion of an initial section on the distinction between *extensional definitions* and *intensional definitions* essential, and then attempt to clarify the definition itself. Subsequently I turn to some of the difficulties that Tarski's definition encounters and explain how these difficulties lead to very innovative solutions. In the final section of the paper, I have chosen to focus on the line of criticism discussing the possibilities in, and limits of, extensionalism, in particular the objections of Hilary Putnam and Simon Blackburn. In addition, I discuss John Etchemendy's and Donald Davidson's reflections on these criticisms.

In my efforts to explain and discuss Tarski's definition of truth, the following books have been an immense help and should be acknowledged: Mark Platts: "Ways of Meaning", Richard Kirkham: "Theories of Truth", Devitt & Sterelny: "Language and Reality", and Collin & Guldmann: "Sprogfilosofi".

It should be noted that my exposition of the objections from Putnam, Etchemendy, Blackburn, Davidson etc. does not do the critics full credit. The objections are laid out in a somewhat simplified manner in this paper, as a means to exposing their inherent difficulties more convincingly.

TARSKI'S SEMANTIC CONCEPTION OF TRUTH

In 1933 the Polish American logician and philosopher Alfred Tarski (1901-1983) published an analysis of truth, later entitled 'The Concept of Truth in Formalized Languages'. The analysis, originally published in Polish, formed the basis of the truth definition subsequently named *The Semantic Conception of Truth* – a truth definition that was to have an enormous impact on contemporary philosophy. Inspired by contemporary logical atomism and positivism, Tarski intended to construct a formally correct definition of truth and thus a definition adequate for application in formal languages. In light of logical positivism, a sound and consistent definition of truth was absolutely necessary. The term *true* was not only an obligatory and essential part of language, but the whole notion of *truth* played a vital role in scientific disciplines and scientific discourse – accordingly in the philosophical discourse. Consequently, Tarski aimed to piece together an unambiguous semantic conception of truth; a truth definition that would pave the way for utilizing the term in the sciences. However, this entailed that Tarski created an absolutely consistent truth definition and furthermore a definition that contained no ambiguous terms, since these would lack applicability within the sciences. To fulfil this ambition, all semantic terms within the definition would somehow have to be reduced or explained in physical terms instead. In Tarski's view the natural sciences demanded a definition of truth that could be understood solely in these natural terms.

As a result, his definition was constructed as an extensional analysis and truth was defined in terms of sentences. According to Tarski, defining truth in terms of sentences had one advantage; the notion of a sentence was to some extent unambiguous. This was to be compared with for instance propositions, statements, and utterances that were, and still are, controversial issues in the philosophy of language[1]. Nevertheless, while defining truth for sentences avoided certain problems, there remained a disadvantage in choosing sentences as the domain for a truth definition: sentences are relative to a particular language and thus any definition of truth would be a notion defined relative to a chosen language. This initial difficulty produced a wide range of problems for Tarski's definition; much of this paper is devoted to the discussion of these.

Tarski was also attentive to the fact that the predicate *true* would be meaningless for a range of sentences, for instance sentences in the imperative mood or questions. Any theory attempting to define truth for sentences must therefore constrain itself to operate with descriptive (or in grammar known as declarative) sentences only.

Now, requirements for any adequate theory of truth depend at large on the method employed and in order to understand Tarski's *Semantic Conception of Truth*, one must be acquainted with the notion of *extensional definitions*.

EXTENSIONAL AND INTENSIONAL DEFINITIONS

Similar to the concepts of *Sinn und Bedeutung,* derived from Frege, we distinguish between the *intension* and the *extension* of words. The *extension* of a word is fairly equivalent to Frege's *Bedeutung* and it is usually conceived as the referent of a particular word; the extension of the word 'book' is every present, previous, and future instance of books.

In addition, an extensional *definition* of a word consists in delineating the set of all the references of the particular word. For instance, defining the term 'countries in Scandinavia' extensionally is done in the following way: (1) determine the different extensions of the term and (2) define the term in relation to the set extension.

> Def. 'Countries in Scandinavia'
> 1. 'Denmark, Sweden, Norway'
> 2. Def. 'Countries in Scandinavia' = {Denmark, Sweden, Norway}

Now, if I set out to define the term *book* extensionally, I would then have to uncover the set extension of the term. This includes uncovering all present, previous, and future tokens of books. In other words, an extensional definition consists in constructing the set of all tokens of the particular term.

However, an obvious weakness in an extensional definition remains: the extensional definition can not yield an unequivocal explanation of the *meaning* of a term, but only provide a set of references of the term's extension. If I ask the question: "What is a book?" – the extensional definition can only yield an ostensive answer, which is more or less satisfactory. But if I ask: "Why is this item a book" – then an extensional definition can not yield any satisfactory answer beyond "it just is!" The underlying problem with the extensional analysis is its failure to reflect that two different terms can have the same reference and yet convey different meanings. Frege's classic example of *'the morning star'* and *'the evening star'*[2] is a lucid illustration of this. Both terms refer to the planet Venus, but they convey different meanings, or have different

---

[1] Harnish p.537, Tarski: The Semantic Conception of Truth
[2] Frege: "On Sense and Reference" 1892, in Harnish 1994.

intensions. Therefore, neither term – *the morning star* or *the evening star* – can be substituted *salva veritate* in every sentence.

If we are looking for an analysis that can satisfy the meaning of a certain term, we must instead turn to intensional definitions. We owe it to Frege to recognize this problem and hence distinguish *Sinn (somewhat equivalent to intension)* from *Bedeutung (somewhat equivalent to extension).*[3]

The *intensional* definition differs from the *extensional* in its effort to explain the *meaning* of a particular term. The notion of adequate intensional definitions is a very complicated and controversial issue within the philosophy of language, but I will attempt a very brief and superficial clarification. Again, if I set out to define the term *book*, this time intensionally, I must account for all the necessary and sufficient conditions for something being a book. Now, constituting a necessary condition for being a book would, for instance, be that the object in question contained pages. This seems a necessary condition for any object claiming to be a book, but this condition alone can not constitute *being a book* (a newspaper contains pages as well, but it is not a book). Besides pages, a book must also contain writing of some kind, symbols (that is, letters of the alphabet), perhaps even a title, a front and back cover and so on. When all the necessary and together sufficient conditions of the particular term are accounted for, we have constructed an intensional definition of the term. Consequently the definition aims at *describing* the term unambiguously.[4]

EXTENSIONALIST THEORY

It is imperative to note that Tarski intends to provide a satisfactory *extensional* definition of truth for a given language, and not an *intensional.* This necessitates that a set extension of all true sentences can be constructed, or that he, in other words, must compose a complete list of all true sentences in order to succeed in adequately defining truth extensionally. However, we can explore his strategy by utilizing the method in a less complicated model: we define truth for a particularly simple language, which we name L_1. We assume that L_1 contains only the two following sentences:

'Grass is green'
'Snow is white'

Both these sentences are obviously true and given our earlier assumption – that these are the *only* sentences in our language - they constitute an extensional definition of truth for the language L_1.

This all seems fairly uncomplicated – now all we need is to compile the list of true sentences for any natural language such as English, German, or French. Unfortunately constructing such a list poses severe difficulties. First of all, natural languages contain more sentences than our previously constructed language L_1. Actually a natural language contains a potential infinity of meaningful sentences, but more on this later. Instead, let's look at some initial problems for an extensional definition of truth. Consider the following sentences,

'Lee Harvey Oswald killed John F. Kennedy'
'Syria will be bombed next week'
'Julius Caesar slept on his stomach on his 18th birthday'

---

[3] cf. "On Sense and Reference" 1892, in Harnish 1994.

[4] This explanation of intension is somewhat derived from Frege's notion of *Sinn.* But it needs mention that Frege's account of necessary and sufficient conditions is a very controversial issue. Many theorists, e.g. Hilary Putnam and Saul Kripke, have put forth strong objections to the notion of necessary and sufficient conditions.

The above sentences are rather uncontroversial sentences in a natural language, but obviously the truth values of either of these sentences are dependent on unknown states of affairs. Seemingly, the truth value of the sentences cannot be determined and thus a *complete* list of true sentences cannot be constructed. This is an intolerable result for an extensional definition of truth and therefore it must be resolved. In response to sentences with unknown truth values, Tarski instead constructs a theorem that specifies the truth condition for any arbitrary sentence in our chosen language.

TRUTH CONDITIONS

What are the requirements for generating a true sentence? According to Tarski, any theory of truth should aim to justify natural intuitions on truth and consequently

> "[..] do justice to the intuitions which adhere to the classical Aristotelian conception of truth – intuitions which find their expression in the well-known words of Aristotle's Metaphysics:
> To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true."[5]

Hence, Tarski must construct a definition of truth where the truth conditions for each and every sentence of the language are logically entailed, and in addition is an intuitively plausible definition. Tarski goes on to say in the following passage:

> "If we wished to adapt ourselves to modern philosophical terminology, we could perhaps express this conception by means of the familiar formula:
> The truth of a sentence consists in its agreement with (or correspondence to) reality."[6]

Consequently, Tarski proceeds in specifying truth conditions: For any sentence in the language, e.g. L_1, the truth condition is the state of affairs that must obtain in order to make the sentence true. Accordingly, the sentence 'snow is white' will be true if in fact snow *is* white. This is the truth condition for the sentence 'snow is white'. This formula specifies the truth conditions of any given arbitrary sentence in our chosen language:

> 'Snow is white' is true if, and only if, snow is white

– and it is formalized into the following theorem,

> T: 'S' is true if, and only if, *p*

The sentence in quotation marks, 'S', represents a *structural* description of the sentence for which we are specifying truth and *p* represents the state of affairs that must obtain in order to make the sentence true[7]. Tarski named this theorem convention-T (it has also been known as *scheme T*, *formula-T*.), and it constitutes a test for *material adequacy*[8]. Now, it is important to note that convention-T is *not* Tarski's definition of truth, but only a *test of the adequacy* of the particular truth definition. By using convention-T we can specify truth conditions for every sentence in our chosen language, and in this way construct an extensional *conception* of truth. It is also evident that any adequate

---

[5] Harnish p.538, Tarski: The Semantic Conception of Truth
[6] Harnish p.538, Tarski: The Semantic Conception of Truth
[7] Understanding the sentence as a structural description – adding the quotation marks – ensures that the theorem avoids redundancy problems.
[8] Harnish p.540, Tarski: The Semantic Conception of Truth

extensional definition of truth must entail T-sentences for every possible sentence of the language.

The test for material adequacy consists in 'S' being a *structural description* of a sentence in the chosen language and p being the equivalent of S – e.g.:

> S^n^o^w^^i^s^^w^h^i^t^e is true if, and only if, snow is white (where ""^"" serves as concatenation)

Or in other words, the meaning of S is structurally captured by p.

The material biconditional ("if and only if" or its logical symbolic equivalent "$\equiv$") constitutes the test for adequacy, given its truth table. In order to obtain, both sentences must be either true or false.

The above semantic conception of truth would be adequate if the number of sentences contained in natural languages was finite. Then every sentence in the chosen language would entail T-sentences and thereby satisfy the material adequacy condition. Let's explore this improved strategy by constructing a new simple language L_2, containing only the five following sentences:

> T: 'Grass is green' is true if, and only if, grass is green
> T: 'Snow is white' is true if, and only if, snow is white
> T: 'Lee Harvey Oswald killed John F. Kennedy' is true if, and only if, Lee Harvey Oswald killed John F. Kennedy
> T: 'Syria will be bombed next week' is true if, and only if, Syria will be bombed next week
> T: 'Julius Caesar slept on his stomach on his 18th birthday' is true if, and only if, Julius Caesar slept on his stomach on his 18th birthday

It is clear that this account of truth solves the problem with the unknown truth values of some sentences, and so it seems to be an adequate conception of truth for L_2.

However, another significant difficulty remains unsolved: natural languages contain innumerable sentences. A competent user of a natural language is capable of constructing, from a finite set of semantic entities, a potential infinity of sentences.

RECURSIVE DEFINITION

Words like *and*, *or*, *if… then* and others are the reason why natural languages contain a potential infinity of meaningful sentences. In logic these are called sentential connectives, and they function as the conjoining element between atomic sentences. In natural languages the connectives enable us to construct compound sentences by combining atomic sentences to develop new – and there are no clear limits to the potential length of any sentence. That is, from two atomic sentences and the connective *and* we could form the new sentence

> 'Snow is white *and* grass is green'

An extensional conception of truth must consequently be proficient to deal with this aspect of natural languages and thus the definition is required to generate truth conditions for any potential sentence in the language. This poses an obvious difficulty with Tarski's definition: how is he to define truth for unknown sentences?

By adding the sentential connectives to the definition; that is, to the extension of potentially true sentences, Tarski constitutes what he calls a *recursive definition* of truth.

The connectives now enable us to construct a potential infinity of sentences from the finite set of semantic entities in the language. But we are also able to derive truth conditions from any potential sentence in the language, by referring to the rules of the connectives and their truth tables. One by one we can divide any compound sentence and generate the truth condition for every atomic sentence; this in turn yields the truth condition for the entire compound sentence. A brief example – the sentence 'snow is white and grass is green' contains two atomic sentences, each true under certain conditions:

> 'Snow is white' is true if, and only if, snow is white.
> 'Grass is green' is true if, and only if, grass is green.

And thus the truth condition for the entire compound sentence is derived:

> 'Snow is white and grass is green' is true if, and only if, snow is white and grass is green.

According to the truth table for the conjunction *and*, both sentences combined by a conjunction must be true to establish truth of the entire compound sentence.
Thus, using our language L_2 we are now capable of constructing a potential infinity of sentences and deriving truth conditions as well:

> 'Snow is white and grass is green and snow is white and grass is green and Syria will be bombed next week and… '.

Any atomic sentence can thus be reused with the aid of connectives in natural languages.

THE ANTINOMY OF THE LIAR

A major challenge in constructing a definition of truth lies in solving the problem of paradoxes in language. Consider the following sentences,

> 'This sentence contains thirty eight letters'
> 'This is an English sentence'

Here we have two very simple sentences that are self-referential, as the very meaning of each sentence concerns the sentence itself. In these particular cases there are no apparent problems for natural languages. But consider the following sentence,

> 'This sentence is false'

The evident problem with this sentence is that it is both true and false at the same time; if the sentence is true, then from its meaning it follows that it is false, and if it is false then from its meaning it follows that it is true. Either way, the sentence will be both true and false and, hence, a paradox.

> S: 'This sentence is false' - yields the following truth condition:
> 'S' is true if and only if 'S' is false

Now, so far Tarski's proposed conception of truth has necessitated the principle of bivalence, or in other words the law of the excluded third: sentences can be only true or false – there is no third option, since a third option would make the extensional conception of truth unattainable.

Tarski asserts that the cause of the before mentioned paradox, *the antinomy of the liar[9],* is two basic assumptions that all natural languages presuppose. The assumptions are:

> "1. We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term '*true*' referring to sentences of this language […] A language with these properties will be called '*semantically closed*.'
> 2. We have assumed that in this language the ordinary laws of logic hold."[10]
> [Tarski's italics]

It is these two assumptions that conflict when paradoxes occur in natural languages. Consequently, one of these assumptions must be rejected if Tarski's theory of truth is to avoid *the antinomy of the liar.* It seems clear that rejecting the second premise, *the ordinary laws of logic,* would be more problematic for any theory of truth, and instead it is the first premise that must be discarded. This compels Tarski to abandon the concept of *semantically closed* languages – natural languages. He creates a distinction between an *object language* (the language for which we are specifying truth) and a *meta-language* (a language in which the semantic terms such as *truth*, *satisfaction,* and *reference* are contained).

This is also the dismissal of natural languages. Tarski's truth definition will only apply to semantically open languages – languages that do not maintain the possibility of self-referring sentences. So if we are dealing with certain reduced constructions of natural languages, where the notion of *truth* and semantic terms have been eliminated, Tarski's definition still may apply. But in general natural languages will not be adequate for the application of Tarski's definition.

DEFINING TRUTH

Until now I have not commented on the crux of Tarski's semantic conception of truth, the notions of *satisfaction* and *reference.* Tarski's method – defining truth through the semantic terms *satisfaction* and *reference* – has been praised by many logicians and formalists and this part of Tarski's definition was undoubtedly very innovative. Not surprisingly, this part of the definition is also somewhat complicated. Hence, explicating the notions of satisfaction and reference is not without its difficulties, and the following paragraph therefore only intends to give a fairly superficial explanation of the terms. I feel the definitions of *satisfaction* and *reference* are of importance if one wishes to understand the central ideas underlying the semantic conception of truth.

SEQUENCES, SATISFACTION, AND REFERENCE

When Tarski commences in explaining the predicate *true* through semantic terms, he is actually trying to resolve yet another difficulty in natural and certain formal languages. When dealing with complicated languages such as natural languages, one encounters so-called open sentences. In natural languages, open sentences are grammatically complete sentences, but they contain variables. In turn they cannot be explained essentially as sentences, because they do not say anything substantial; an open sentence does not contain substantial information until it is somehow closed and, hence, open sentences are neither true nor false. In order to close an open sentence one must either affix a name or object for each variable in the sentence or affix a quantifier for each variable. Two examples of open sentences:

---

[9] Harnish p.542, Tarski: The Semantic Conception of Truth: Tarski names the paradox *the antinomy of the liar.*
[10] Harnish p.544, Tarski: "The Semantic Conception of Truth", 1944

1. 'x is black'
2. 'x is the teacher of y'

- affixed with names, the sentences are closed:

1. The book is black
2. Socrates is the teacher of Plato

- or affixed with quantifiers, the sentences can be closed:

1'. $(\exists x)(x$ is black$)$
2'. $(\exists x)(\exists y)(x$ is the teacher of y$)$

These sentences are grammatically correct and when closed both sentences also contain substantial information – as a result they both have a truth value. One then substitutes a variable for an object to attain the truth value of the sentence. However, with this a well-known problem resurfaces. Earlier I elaborated on how natural languages contain a potential infinity of sentences due to sentential connectives, and this problem remains: for any given predicate and quantifier a potential infinity of open sentences can be closed, as the example below should illustrate:

$(\exists x)(\exists y)(\exists z) \ldots (x$ is black and y is black and z is black$) \ldots$

Now, the difficulty remains; it seems that no obvious method exists that can explain the truth value of a potential infinity of quantified sentences. Tarski's stroke of genius was not only to create a way of assigning truth to quantified sentences, but to construct a recursive method for this problem as well. Tarski was aware that in order to determine truth for any sentence containing quantifiers and variables we would need to understand that the variables are truth-functional operators. The objects we assign to the variables in a given sentence will determine the truth value of the sentence. On the basis of this insight, Tarski asserts that a sentence containing quantifiers is true if and only if it *all objects in a sequence* are satisfied. This is in essence Tarski's definition of truth. Now, this may seem rather unclear and obviously the notions of *satisfaction* and *sequences of objects* must be explicated.

Satisfaction: Suppose we are dealing with an object 'the car' and the open sentence 'x is black', then 'the car' will satisfy the predicate 'x is black' when the car is black. *Satisfaction* is defined through set theory – if the object, x, is a member of the predicate set, then the object is satisfy the predicate.

Tarski's notion of *sequences* can be understood as a test of objects satisfying the predicates in quantified sentences. Dealing with a sentence containing more than one variable we must, correspondingly, employ more complicated sequences in order to determine which objects satisfy the different predicates in the sentence:

'$(\exists x_1)(\exists x_2)(\exists x_3)(x_1$ is the teacher of x_2 and x_1 was student of $x_3)$'

For every variable in the sentence an object must be substituted. Hence we construct a sequence of objects and the truth value of the sentence will then be obvious. Any object can be inserted into the sequence ensuring that all potential interpretations of the variables are properly analysed. Consider the following example where we are dealing with a sequence of objects as the following:

x_1: Plato, x_2: Aristotle, x_3: Socrates ….

In this case the predicates are satisfied by the objects and hence the entire sentence is true. Whereas, if we are dealing with a sequence as the following:

x_1: Aristotle, x_2: Socrates, x_3: Plato….

- Then the predicates are not satisfied by the objects and the sentence is thus not true. The predicates in quantified sentences are satisfied as according to an extensional definition of objects within the set of the predicate-domain. Objects satisfied by the predicate 'the teacher of Plato' must thus be included in the extensional definition – the set of teachers of Plato. Furthermore, the order of the objects in a sequence will also be decisive for the truth value of a quantified sentence. A closed sentence containing multiple variables must therefore not only have its variables satisfied as according to the extension, but have the correct order in the sequence as well. Consider the following sentence

'x_2 is white, and x_1 is white, and x_2 is frozen water'

We test the following sequence of objects:

x_1: a swan, x_2: snow

In this case the sentence is true, but one notices that the order of the sequence is essential to the truth value. Both names 'swan' and 'snow' satisfy the set of 'white objects' but the sequence order ensures the correct interpretation, since 'swan' does not satisfy the predicate *is frozen water*.

If Tarski's semantic conception of truth is to be successful, then every predicate and name of the language we are using must be properly defined as according to set theory. Also *reference* is explained as according to membership in certain sets, and thus without containing any semantic terms.
This is why Tarski himself proclaimed that his theory of truth would not apply to natural languages. The predominant problem is that natural languages are semantically closed languages, but problems of ambiguity also pose an immense difficulty. Ambiguities must be removed if Tarski's definition is to be consistent and it seems highly unlikely that this can be achieved.

Nonetheless, the American philosopher Donald Davidson[11], among others, has made an exceptional effort to use Tarski's truth definition in theories of meaning. Davidson insisted on the compatibility of Tarski's definition and natural languages.

RECEPTION AND MISTAKEN OBJECTIONS

The reception of Tarski's definition of truth has been unusually diverse; from Karl R. Popper embracing Tarski's theory as vindicating *the free use of the intuitive idea of truth as correspondence to facts[12]*, to critics like Hilary Putnam, who blatantly ridiculed Tarski's definition as an absolute failure.

Tarski's semantic conception of truth has been extensively discussed within academic circles and a wide range of objections to his truth definition has been put forth.

---

[11] cf. Davidson's early work: "Truth and Meaning" (1967) and "In Defence of Convention T" (1973). But also later work, e.g. "The Structure and Content of Truth" (1990) promotes a Tarskian style of semantics, although Davidson at this point revises his position on truth theories somewhat.
[12] Susan Haack: "Philosophy of Logics" 1978, Cambridge University Press p.112

However, several objections have been downright misunderstandings and misinterpretations of Tarski's work. For this reason, the broad discussions of Tarski often contain faulty and problematic objections that in essence do more harm than good[13]. One very common misunderstanding is the objection that Tarski never explains how to uncover whether or not a specific state of affairs in fact obtains. This objection fails to recognize that Tarski's definition doesn't aim to answer that question. Tarski shows how to construct a sound and consistent definition of truth, but he never intends to answer the question of how we confirm the truth of any sentence in the language. Knowing what will make a sentence true and knowing how to uncover the truth of that sentence, are in essence two very different things.

I will start by addressing an analysis of Tarski's definition proposed by Susan Haack in 1978 and then proceed to a discussion of objections from critics such as Hilary Putnam, Simon Blackburn, John Etchemendy, and Donald Davidson.

SUSAN HAACK: MISUNDERSTANDING TARSKI

Considering Susan Haack's analysis of Tarski's definition in 'Philosophy of Logics'[14], one will encounter a somewhat radical interpretation of Tarski's work. Haack claims that convention-T is an inadequate test of the truth definition. This objection, however, is in fact fairly misleading. Haack is aware that convention-T is *not* Tarski's *definition* of truth and emphasizes that Tarski on this point has been somewhat misconstrued by certain critics, but she nevertheless asserts that convention-T remains inadequate. The reason for this is that convention-T seems to be potentially consistent when combined with a number of very bizarre truth definitions. Haack writes:

> "Consider the following definition of truth, which seems to me definitely bizarre: a sentence is true iff it is asserted in the Bible. [..] (I shall call it [the definition] 'D$_B$' for short)
> [..] Now it is indeed the case that someone who did not accept D$_B$ might deny:
> 'Warsaw was bombed in World War II' is asserted in the Bible iff Warsaw was bombed in World War II.
> But further reflection makes it clear that a proponent of D$_B$ could perfectly well maintain that his definition does entail all instances of (T) [..]"[15]

In my opinion, this objection constitutes a minor misunderstanding. Tarski insisted on convention-T as the *entailment* of any adequate truth definition and as Richard Kirkham[16] somewhat correctly points out in "Theories of Truth", Haack's proposed definition ("a sentence is true iff it is asserted in the Bible") does not *entail* all T-sentences. It is only *compatible* with T-sentences and consequently, it is not an adequate definition of truth. However, she could argue (and probably would) that certain very religious people might find the definition intuitively plausible, and accordingly the definition would, in their view, entail T-sentences. Conversely, the consequence of this argument is the contention that natural intuitions cannot constitute any basis of a definition, because intuitions in essence differ among individuals[17]. This line of argument, in my opinion, fails to understand Tarski's intention – which exactly was to define truth extensionally and thereby catch hold of our natural intuitions of truth.

---

[13] cf. Kirkham: "Theories of Truth" 1992, MIT Press.

[14] Susan Haack: "Philosophy of Logics" 1978, Cambridge University Press

[15] Susan Haack: "Philosophy of Logics" 1978, Cambridge University Press p.101

[16] Richard Kirkham: "Theories of Truth" 1992, MIT Press p. 175

[17] Kirkham determines Haack's objection as a misunderstanding, but I feel that his explanation is somewhat incomplete. Haack's claim is a radical one if it is to be sound, mainly the claim that intuitions are useless when defining concepts. But one could assert such an argument and thus pose a sound objection. This would not be a misunderstanding, but merely a very radical claim.

Furthermore, the above line of argument determines all extensional analysis as simple stipulations, which seems to be an oversimplification of extensionalism.

Now, there are other more advanced and reasonable objections to Tarski's definition that also concern the relation between natural intuitions and stipulations, whereof some of the more significant ones will be discussed in the following sections.

HILARY PUTNAM: TRUTHS OF LOGIC

Few, if any, have so blatantly ridiculed the semantic conception of truth as American philosopher Hilary Putnam. Putnam raised his objection against Tarski in a short article entitled 'A Comparison of Something with Something Else'[18] – an article based on a lecture on Willard Quine and Richard Rorty. The argument, although significant, was in fact fairly uncomplicated. Putnam merely asserted that Tarski's definition results in a stipulated account of truth – a definition of truth not necessarily correlated with reality nor in agreement with the use of language. According to Putnam, the cause of this defect was Tarski's method. Progressing as Tarski did – eliminating all semantic terms within the object language and constructing an extensional recursive definition – merely resulted in a list of logical truths.

> "(2) (for any sentence X) If X is spelled S-N-O-W-SPACE-I-S-SPACE-W-H-I-T-E, then X is true in L if and only if snow is white.
> Now pay close attention, please! This is just where, it seems to me, philosophers have been asleep at the opera for a long time! Since (2) is a *theorem of logic* in meta-L (if we accept the definition – given by Tarski – of "true-in-L"), since no axioms are needed for the proof of (2) except axioms of logic and axioms about spelling, (2) holds in all possible worlds."[19]

Putnam asserts that the T-sentence "'snow is white' is true if, and only if, snow is white" is logically true and the apparent problem is that the truth of 'snow is white' thus seems to be dependant on logic as opposed to contingent facts. However, the result of any definition will to a certain degree be a stipulation and in the case of Tarski's truth definition the end result does consist of *provable theorems*, as Putnam writes. Consequently, a problem with Tarski's definition seemingly remains unsolved according to Putnam. Nevertheless, Tarski often emphasized the importance of a truth definition meeting the demand of logically entailing convention-T, and thereby producing the proper truth conditions for any statement in the language. Tarski was well aware of the limits of definitions and he therefore regularly stressed that defining truth was only possible relative to a specific language. His extensional recursive definition was an attempt to do just this and nothing more. But, according to Putnam, the definition – being either a list of true sentences or a recursive definition, remains inadequate, because the result of an extensional analysis does not reflect linguistic use. *True-in-L* essentially develops into *true-in-any-possible-world* and in addition *true-in-L* becomes potentially detached from actual use of language. This can be illustrated by considering the implications of a logical truth: assuming that the truth of the sentence 'snow is white' is not dependant on contingent facts or on *use*, but solely dependant on the definition. This would mean the following: in a possible world, when a speaker utters the sentence 'snow is white', she is uttering a true sentence, even though the words uttered might actually mean *snow is green*. Putnam thus concludes that:

---

[18] Hilary Putnam: "A Comparison of Something with Something Else" in "NLH" 1985, vol.17
[19] Hilary Putnam: "A Comparison of Something with Something Else" in "NLH" 1988, p.63

> "As a philosophical account of truth, Tarski's theory fails as badly as it is possible for an account to fail."[20]

The claim made by Putnam is by no means insignificant. It touches on a profound difficulty for not only extensional analysis, but also definitions in general. However, in my opinion, Putnam's objection does not pose a substantial problem for Tarski. One must take into account that Tarski never set out to define a trans-linguistic concept of truth. In fact, Tarski repeatedly emphasized that truth could only be defined relative to a specific language and thus be nothing more than a stipulated account approximating natural intuitions on truth. The result – a recursive definition of truth providing truth conditions for any potential sentence of the language – is thus to some extent a stipulated account of truth. But, as Tarski writes:

> "The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion. We must then characterize this notion precisely enough to enable anyone to determine whether the definition actually fulfils its task."[21]

I suspect that Putnam just plainly disagrees that Tarski's definition succeeds in doing this.

SIMON BLACKBURN: STIPULATIVE NOTIONS

Along the same line of argument as that of Putnam, Simon Blackburn in his book 'Spreading the Word' put forward an objection to Tarski's definition. Equivalent to Putnam's claim, Blackburn also objects that Tarski's definition merely establishes a stipulation of truth. Blackburn emphasizes the extensional analysis used in defining semantic terms such as *satisfaction*, *reference,* and *truth* for specific languages as an inadequate approach. The result, the extensional definition, merely yields a stipulated account of different semantic terms delimited to a specific language. Nevertheless, Blackburn presents his objection in a way somewhat different from Putnam's. By considering the use of sentences in a specific language, Blackburn remarks that uttering a meaningful sentence can not usually be done without discerning the semantic implications of the used sentence[22]. But in view of Tarski's definition, the sentences in the language are true as according to the extension – the list of true sentences – and can thus be determined as true or not true by a user of the language disregarding the semantic properties of the sentence.

Moreover, Blackburn sheds light on another problem with Tarski's definition. Blackburn considers the relation between different definitions of truth. Let's consider two different languages, L_1 and L_2, each having a distinctive definition of truth. We presuppose two very simple languages, and consequently define truth in the following more or less simplified way:

> L_1 contains the sentence S: 'Snow is white'
> S is true-in-L_1 if, and only if: S = 'Snow is white' and snow is white
> L_2 contains the sentence S: 'Grass is green'
> S is true-in-L_2 if, and only if: S = 'Grass is green' and grass is green

---

[20] Hilary Putnam: "A Comparison of Something with Something Else" in "NLH" 1988, p.63
[21] Harnish p.537: Tarski: "The Semantic Conception of Truth"
[22] Simon Blackburn: "Spreading the Word" 1984, p.266

In order to accord with Tarski, we assume that both the names and predicates of these sentences are properly defined by means of satisfaction and reference, and thus we maintain L_1 and L_2 as adequate definitions of truth. Now we have only to refer to the definition (the lists above) to determine truth in either of the languages. But, according to Simon Blackburn, this does not reflect any pre-theoretical definition of truth – it actually presupposes it. This can be demonstrated by comparing the languages L_1 and L_2. One then realises that the definition of truth assigned to the language L_1 has nothing in common with the definition of truth assigned to the language L_2[23]. Truth is *defined* for each language and consequently the truth definition is completely exclusive to the specific language.

This means that even though we can intuitively accept the definition as given by Tarski, it seems rather strange that two different definitions of truth, that both seem intuitively plausible, have absolutely nothing in common whatsoever.

Blackburn then asks what would happen if we were to add a new sentence to L_1, e.g. the sentence, 'The earth is round'. This would cause no immediate problem. We would simply adjoin the sentence to the extension in the following way

> L1 contains the following sentences:
> S1: 'Snow is white' and S2: 'The earth is round'
> S is true-in-L1 if, and only if: S = (S1 = 'Snow is white' and snow is white) or
> (S2 = 'The earth is round' and the earth is round)

But the question is why we are able to carry out this addition of a new sentence? Blackburn claims that a concept of truth must be presupposed – if it weren't, we would have absolutely no idea how to conjoin a new sentence to an existing extension. How would we for example conjoin the sentence 'КИМ ЖД ІЦ ЦФЪЫ' into the extension of truth? Obviously this would be impossible, since in this case we have no idea what the sentence means, and hence the state of affairs that must obtain in order to make the sentence true, is unattainable. Consequently, it is clear that Tarski has not only presupposed meaning in his theory, but he has presupposed an ability to identify truth as well.

The objections cannot be easily refuted. One line of argument that might defend Tarski's definition could be emphasizing the limits of an extensional analysis, but Blackburn's objection goes beyond this. The main problem remains the following: suppose Tarski had constructed specific extensional definitions of truth for all natural languages; it would then seem easy to define the notion of trans-linguistic truth. We simply adjoin all the definitions of truth relative to specific languages in a general extension named *trans-linguistic truth*. But this is not possible, and this is the gist of Blackburn's objection. Providing a general account of truth – an extensional definition of trans-linguistic truth by accumulating the different truth definitions – would create paradoxes in the extension and hence be inconsistent. Tarski can only provide relative definitions that are dependant on additional meta-levels of language. Combining all the truth definitions in one extension will consequently become internally inconsistent. One must admit that a link between the different truth definitions seems to be missing. If Tarski's definition does capture an intuitive notion of truth given convention-T, then the problem remains that we cannot give a general account of truth for all definitions.

---

[23] Simon Blackburn: "Spreading the Word" 1984, p.267

JOHN ETCHEMENDY: RECOGNISING TARSKI'S AIM

In 1988 John Etchemendy wrote a short article entitled 'Tarski on Truth and Logical Consequence' that addressed Putnam's objection in particular. Etchemendy argues that Putnam's objection rests on a mistaken presupposition – mainly the assumption that Tarski's work somehow aims at providing a pre-theoretical notion of truth, and in addition that a stipulative definition of truth is intrinsically problematic. Etchemendy maintains that the term *true,* according to Tarski, is essential to everyday language and to scientific discourse. Tarski merely aimed at constructing a sound definition with a main concern for adhering natural intuitions and consistency – and not, as Putnam and Blackburn seem to presuppose – a concern for semantic implications. Etchemendy states:

> "[..] achieving this goal does not demand an explicit definition of truth. [..] Tarski entertains various other approaches to the problem, such as treating the notion of truth as primitive and giving it an axiomatic characterization. Introducing the concept of truth into a theory by means of an explicit eliminative definition is just one way to meet the challenge."[24]

The presupposition that Tarski's truth definition must illuminate the semantic properties of natural languages is a misunderstanding, according to Etchemendy. Although he agrees with Putnam and Blackburn that Tarski's definition reduces *truth* to logic, syntax, or set theory, he also argues that the lacks of semantic implications are unproblematic. According to Etchemendy, this was exactly what Tarski intended – a stipulative account of truth:

> "For Tarski's aim was to eliminate the semantic notions, notions whose consistency seemed questionable, in favour of syntactic and set-theoretic notions."[25]

Therefore, asserting the objections of Putnam and Blackburn remains a misunderstanding of Tarski's aim. There are no intrinsic problems with stipulated definitions, according to Etchemendy, on the contrary – this is exactly the essence of definitions; they are not dependent on any contingent facts or intuitions. So obviously, if we try to attain a theory of meaning from the machinery established by Tarski, we will encounter problems and consequently Putnam's objection remains unsound. And although Tarski aimed at constructing a definition of truth that would reflect natural intuitions on truth – manifested through convention-T, Tarski never intended his definition to redeem any semantic questions.

DONALD DAVIDSON: EMPIRICAL CONVENTION-T

In 1967, American philosopher Donald Davidson published a very important and influential paper entitled 'Truth and Meaning' that constituted the foundation of a new field of semantics. 'Truth and Meaning' concentrates on the question of constructing a theory of meaning and an essential part of the theory rests on the formal machinery constructed by Tarski. Subsequently, Davidson continued to defend a *Tarskian* style of semantics and also insisted on the compatibility between a *Tarskian* conception of truth and a correspondence theory of truth[26]. However, I will not attempt to give a comprehensive account of Davidson's early programme in formal semantics, as that would surpass the ambitions of this paper. Davidson's theory is exceedingly complicated and thus almost impossible to explicate in a few paragraphs.

---

[24] Etchemendy: "Tarski on Truth and Logical Consequence" 1988, p.53-54
[25] Etchemendy: "Tarski on Truth and Logical Consequence" 1988, p. 57
[26] cf. "True to the Facts" (1969) reprinted in "Inquiries into Truth and Interpretation" (2001)

Instead I intend to focus on Davidson's later paper, "The Structure and Content of Truth" from 1990, a paper that among other issues addresses the discussion between Putnam and Etchemendy. Here Davidson abandons the insistence on a correspondence theory of truth and engages in a discussion of plausible alternatives.

Davidson evaluates both the objections of Hilary Putnam (and implicitly Simon Blackburn) and the reply from John Etchemendy – but neither accounts of Tarski's definition seem satisfying. Davidson actually agrees with Etchemendy that Putnam's objection is flawed. Putnam's objection rests on demanding something from Tarski's definition, which Tarski never offered nor stated that the definition would provide. But the focal point of Davidson's discontent with Putnam's argument is in essence the claim that Tarski's definition is rigorously analytic and thus contains no empirical content. In addition, this is also the reason why Davidson is inclined to dismiss Etchemendy's defence of Tarski's definition. Etchemendy asserts, as I have previously explained, that Tarski's definition *is* a stipulation, but also that this constitutes no difficulty for Tarski. But, obviously, if one intends to read empirical content into convention-T as Davidson does, then Etchemendy's defence will not work. Although Etchemendy to some extent acknowledges that Tarski aimed at constructing a definition that would adhere to natural intuitions of truth and that convention-T was the proposed device for this purpose – he still insists on the definition ultimately being a stipulation. So in fact – even though Etchemendy criticises Putnam for rejecting Tarski on incorrect grounds, they in fact agree that the definition consists of logical truths. The disagreement rests on whether or not this poses a real problem for Tarski. To this Davidson states:

> "[..] we recognize as true all sentences of the form " 'Snow is white' is true if and only if snow is white." Tarski calls such sentences "partial definitions" of truth. Obviously, a definition that entails all such sentences will have the same extension as the intuitive concept of truth with which we started. To admit this is to count T-sentences as having empirical content; otherwise convention-T would have no point. [..]"[27]

In my opinion, Davidson here makes an important and decisive point. Putnam's objection and Etchemendy's reply both recognize the extensional definition of truth as strictly stipulative, but this is a mistake. Davidson's contention is in essence that constructing any extensional definition without due regard to empirical groundings are indeed stipulative, but such definitions are also worthless:

> "Suppose we offer as a definition of the predicate 'x is a solar planet' the following: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto. This entails the P-sentence 'Neptune is a solar planet'. Is this last a logical truth? One may well say so if our definition is purely stipulative, otherwise not. The question whether it is purely stipulative is not one that can be answered by studying the formal system; it concerns the intentions of the person making the definition."[28]

There is a reason why convention-T appears very plausible; namely that convention-T reflects our natural intuitions. Davidson's thus argues that this is the case only as convention-T is endowed with an empirical content. Davidson's line of reasoning, in my opinion, reveals a serious weakness in Putnam's objection and Etchemendy's reply. Furthermore, Davidson's interpretation of Tarski's system yields another

[27] Davidson: "The Structure and Content of Truth" in "The Journal of Philosophy no.6" 1990, p.292
[28] Davidson: "The Structure and Content of Truth" in "The Journal of Philosophy no.6" 1990, p.293

advantage; interpreting Tarski's definition empirically can form the foundation of a truth definition that contains a substantial semantic content. However, Davidson recognises the objection of Blackburn – that the different relative truth definitions seemingly have nothing in common, and that this poses an immediate problem, but Davidson insists that this is a difficulty that will be solved:

> "The concept of truth has essential connections with the concepts of belief and meaning, but these connections are untouched by Tarski's work. It is here that we should expect to uncover what we miss in Tarski's characterizations of truth predicates."[29]

– mainly the characteristics that the different definitions share.

In conclusion, I would like to remark that whether or not one shares the views of Davidson regarding the use of Tarski's formal machinery in semantics, I nevertheless find his argument against Putnam, Etchemendy and implicitly Blackburn very compelling. Although both Putnam and Blackburn state potent objections against Tarski's definition, it seems to me that they are based on something resembling a misunderstanding. Or as for Putnam, his objection presupposes certain assumptions which cannot be found in Tarski's definition and although Putnam's objection touches on a weakness in extensional definitions, it is hardly decisive. Putnam's objection, if anything, illustrates a reluctance to appreciate the purpose of a *definition*. In light of Davidson's argument on empirical content in extensional definitions, I am inclined to think that Blackburn would revise his criticism of Tarski. The fundamental problem for Blackburn was the truth definition being exclusive to the particular language for which it was defined, but understanding the definition as a result of empirical implications, the different definitions may have something in common after all. Etchemendy's defence of Tarski's definition and of stipulative definitions does to some extent aim to recognize Tarski's intention, but I find his conclusion discomforting. Ultimately Etchemendy claims that Tarski merely stipulates his way out of the problems of semantic paradoxes and although this may yield a consistent definition of truth, it does not capture why Tarski's definition seems plausible.

As for Davidson, I agree with his criticism of Putnam and Etchemendy. Furthermore his assertion that convention-T yields empirical content is a convincing argument. It succeeds in recognizing that extensional definitions are more than mere stipulations. A definition may have an end result of stipulated notions, but it aims at catching hold of something intuitively plausible. However, I will not attempt to answer, whether or not Tarski's definition of truth can constitute the basis of semantic theory. In reply to this question Davidson would obviously answer a loud and clear "yes".

---

[29] Davidson: "The Structure and Content of Truth" in "The Journal of Philosophy no.6" 1990, p.295

SYLLABUS AND ADDITIONAL LITERATURE:

1994: 'Basic Topics in the Philosophy of Language' (BTPL),
- edited by Robert M. Harnish

Alfred Tarski:
*1944: The Semantic Conception of Truth and the Foundations of Semantics
– reprinted in BTPL (1994)

Simon Blackburn:
*1984: Spreading the Word, Groundings in the Philosophy of Language
Clarendon Press, Oxford, 1984

Finn Collin & Finn Guldmann:
1998: Sprogfilosofi
Gyldendal, Nordiske Forlag A/S

Donald Davidson:
1967: Truth and Meaning
Reprinted, in 'Inquiries into Truth and Interpretation' 2001
1969: True to the Facts
Reprinted, in 'Inquiries into Truth and Interpretation' 2001
*1973: In Defence of Convention-T
Reprinted, in 'Inquiries into Truth and Interpretation' 2001
*1990: The Structure and Content of Truth
Journal of Philosophy vol.87, 1990
*1996: The Folly of Trying to Define Truth
Journal of Philosophy vol.93, 1996

Michael Devitt and Kim Sterelny:
1999: Language and Reality
Blackwell Publishers, 1999

John Etchemendy:
*1988: Tarski on Truth and Logical Consequence
The Journal of Symbolic Logic vol. 53, 1988

G. Frege:
1892: On Sense and Reference
Reprinted in BTPL, 1994

Susan Haack:
*1978: Philosophy of Logics
Cambridge University Press, 1978

Richard Kirkham:
1995: Theories of Truth
The MIT Press, 1995

S. Kripke:
1980: Naming and Necessity
Reprinted in BTPL, 1994

Mark Platts:
1979: Ways of Meaning
The MIT Press, edition from 1997

Hilary Putnam:
1975: The Meaning of 'Meaning'
Reprinted in BTPL, 1994
1981: Reason, Truth and History
Cambridge University Press, 1981
*1985: A Comparison of Something with Something Else
New Literary History vol.17 (NHL), 1985

NB! Papers marked with '*' are part of the official syllabus.