Tarski's result on the undefinability of truth

Canonical name	TarskisResultOnTheUndefinabilityOfTruth
Date of creation	2013-03-22 13:49:19
Last modified on	2013-03-22 13:49:19
Owner	mathcam (2727)
Last modified by	mathcam (2727)
Numerical id	14
Author	mathcam (2727)
Entry type	Theorem
Classification	msc 03B99
Related topic	IFLogic

Assume **L** is a logic which is under contradictory negation and has the usual truth-functional connectives. Assume also that **L** has a notion of formula with one variable and of substitution. Assume that T is a theory of **L** in which we can define surrogates for formulae of **L**, and in which all true instances of the substitution relation and the truth-functional connective relations are provable. We show that either T is inconsistent or T can't be augmented with a truth predicate **True** for which the following T-schema holds

$$\mathbf{True}('\phi') \leftrightarrow \phi$$

Assume that the formulae with one variable of **L** have been indexed by some suitable set that is representable in T (otherwise the predicate **True** would be next to useless, since if there's no way to speak of sentences of a logic, there's little hope to define a truth-predicate for it). Denote the i:th element in this indexing by B_i. Consider now the following open formula with one variable

$$\mathbf{Liar}(x) = \neg\mathbf{True}(B_x(x))$$

Now, since **Liar** is an open formula with one free variable it's indexed by some i. Now consider the sentence $\mathbf{Liar}(i)$. From the T-schema we know that

$$\mathbf{True}(\mathbf{Liar}(i)) \leftrightarrow \mathbf{Liar}(\mathbf{i})$$

and by the definition of **Liar** and the fact that i is the of $\mathbf{Liar}(x)$ we have

$$\mathbf{True}(\mathbf{Liar}(i)) \leftrightarrow \neg\mathbf{True}(\mathbf{Liar}(\mathbf{i}))$$

which clearly is absurd. Thus there can't be an of T with a predicate **Truth** for which the T-schema holds.

We have made several assumptions on the logic **L** which are crucial in order for this proof to go through. The most important is that **L** is closed under contradictory negation. There are logics which allow truth-predicates, but these are not usually closed under contradictory negation (so that it's possible that $\mathbf{True}(\mathbf{Liar}(i))$ is neither true nor false). These logics usually have stronger notions of negation, so that a sentence $\neg P$ says *more* than just that P is not true, and the proposition that P is simply not true is not expressible.

An example of a logic for which Tarski's undefinability result does not hold is the so-called Independence Friendly logic, the semantics of which is based on game theory and which allows various generalised quantifiers (the Henkin branching quantifier, etc.) to be used.