Properties characterising truth and satisfaction

Bartosz Wcisło

August 30, 2021

Truth theories are obtained by adding to a fixed arithmetical theory (say, $\mathrm{I}\Delta_0 + \exp$ or PA) a fresh predicate $T(x)$ with the intended reading "x is a (code of a) true sentence" and axioms guaranteeing a truth-like behaviour of the new predicate.

One of the most basic kinds of axioms we might add are variants of the Tarski scheme:

$$T\ulcorner\phi\urcorner \equiv \phi, \text{where } \phi \text{ is an } \mathscr{L}_{\mathrm{PA}}\text{-sentence.}$$

If we add the above scheme to PA, we call the resulting theory TB^- ("Tarski Biconditionals"). We call it TB if we also add the full induction scheme for the extended language. These theories also have uniform variants, called UTB^- and UTB:

$$\forall t_1, \ldots, t_n \in \mathrm{ClTerm}_{\mathscr{L}_{\mathrm{PA}}}\ T\phi(t_1, \ldots, t_n) \equiv \phi\big(\mathrm{val}(t_1), \ldots, \mathrm{val}(t_n)\big),$$

where ϕ is an $\mathscr{L}_{\mathrm{PA}}$-formula. Above, $x \in \mathrm{ClTerm}_{\mathscr{L}_{\mathrm{PA}}}$ is an arithmetical formula expressing "x is (a code of) a closed arithmetical term" and $\mathrm{val}(t)$ is the formally computed value of that term (so we also implicitly quantify over these values and we use formalised substitutions which we suppress in order not to clutter the notation).

The theories UTB^- and UTB have very close connection with the notion of (partial, partial inductive) satisfaction classes investigated in the theory of models of PA. In particular, if $M \models \mathrm{UTB}$, then for every $a \in M$, the set of the pairs (ϕ, α) such that ϕ is a standard arithmetical formula, α is a ϕ-assignment with values $\leq a$ (i.e., a function whose domain contains the free variables of ϕ), and $M \models \phi[\alpha]$, is coded in the arithmetical part of M. There is a number of other model-theoretic properties in this spirit:

- If $M \subseteq N$ is a submodel and both M and N are models of UTB^-, then M^* is an elementary submodel of N^*, where M^* and N^* are the reducts of the respective models to the arithmetical language.

- If $M \models$ UTB, then M^* is recursively saturated.
- If $M \models$ UTB$^-$, then Skolem functions are uniformly definable in M. That is, for all $b \in M$, there exists a partial function $F : M^2 \to M$ definable in M such that for each (standard) arithmetical formula $\phi(\bar{y}, x)$ and each tuple $\bar{a}$ of elements smaller than b, if $M \models \exists x \phi[\bar{a}]$, then $F(\phi, \bar{a})$ is a witness for ϕ and the parameters $\bar{a}$.
- If $M \models$ UTB$^-$, then the definability relation is definable. That is, there exists a set $D \subset M^2$ definable in M such that if $\phi(x)$ is a (standard) arithmetical formula with one free variable, then for every $a \in M$, $(\phi, a) \in D$ iff a is the only element satisfying ϕ.

The introduced properties may also typically have "uniform" or "non-uniform" analogues. For instance, instead of considering elementarity between models, we can consider the elementary equivalence relation.

All the listed properties can be considered in the context of some arbitrary theory U extending I$\Delta_0 + \exp$. Most of them seem very closely related to the notion of truth. In our talk, we would like to understand whether this link can be make precise. We try to answer questions of the form:

- If a theory U has a given truth-like property, does it define a truth predicate?
- If a theory U has a given truth-like property P_1 does it also have some other truth-like property P_2?

Some particular instances of these questions have been already considered in the literature. Most notably, Roman Kossak has shown that for every theory U in a countable language which contains PA$^-$ and the full induction scheme for its full language, if for every model of U its arithmetical part is recursively saturated, then in every model of U, one can define a predicate satisfying UTB. We would like to treat the problem of "invariant" characterisations of truth in a systematic manner.

As we already mentioned, there is a number of actual questions behind the two general problems we consider, since we can ask about the definability of truth predicates satisfying different theories, like TB and UTB$^-$, we can ask about various forms of definability (in a model or in the theory U), or we can restrict our attention to different base theories (for instance, the link between Skolem functions and definability is clear in the case of PA, but not in the case of weaker theories). Finally, we can impose further requirements on the theory U (for instance, that it is recursive, or that it is formulated in a finite language).

In our talk, we will try to give a systematic account of the answers obtained concerning the listed questions and mention some of the most interesting or most typical arguments occurring in this line of research. The results described in the talk are a joint work with Mateusz Łełyk.