Gödeli meeldetuletus

Tarmo Uustalu

Teooriapäevad Arulas, 3.–5.2.2003

- **What is this about?** (Rich) languages with a decided intended interpretation, (powerful) theories in such languages, axiomatized (powerful) theories in such languages.

- **Definition:** A *language* L is a first-order logical language with equality and a denumerable amount of non-logical individual, function and predicate symbols. We assume a fixed intended interpretation. This singles out a subset of all L-sentences, the set of *true* sentences.

  $\models A$ means A is true in the intended interpretation.

  An L-*theory* T is a subset of all L-sentences, these sentences are called T-*theorems*.

  $\vdash_T A$ means A is a T-theorem.

  An *axiomatized* L-*theory* is a L-theory generated by a p.r. subset of all L-sentences (called *axioms*) and the inference rules of first-order logic.

- **Definition:** Let T be a theory in a language L (with fixed intended interpretation).
  - T is said to be *consistent* (kooskõlaline), if $\vdash_T A$ implies $\nvdash_T \neg A$ (there are no more theorems than syntactically ok).
  - T is said to be *sound* (korrektne), if $\vdash_T A$ implies $\models A$ (there are no more theorems than semantically ok).
  - T is said to be *syntactically complete* (süntaktiliselt täielik), if $\nvdash_T A$ implies $\vdash_T \neg A$ (there are no less theorems than syntactically ok).
  - T is said to be *semantically complete* (semantiliselt täielik), if $\nvdash_T A$ implies $\not\models A$ (there are no less theorems than semantically ok).

- **Observation:** The semantic properties are stronger than the syntactic ones:
  - soundness implies consistency,
  - and semantic completeness implies syntactic completeness.
- **Observation:** The converses don't hold in general, but:
  - consistency implies soundness under the assumption of semantic completeness,
  - and syntactic completeness implies semantic completeness under the assumption of soundness.
- T syntactically perfect, if it's both consistent and syntactically complete, for then, for every sentence A, either $\vdash_T A$ or $\vdash_T \neg A$ (which mimicks bivalence).
- T is semantically perfect, if it's both sound and semantically complete, for then theoremhood exactly captures truth.

- **Definition:** A language L is *rich* if natural numbers, p.r. operations on natural numbers and p.r. relations on natural numbers are effectively *represented* (esitatud) (faithfully wrt. the intended interpretation) in L by terms, schematic terms and schematics sentences.

  Terms representing natural numbers are called *numerals*.

- **Definition:** An L-theory T is *powerful*, if natural numbers, p.r. operations and p.r. relations on them satisfy the following *presentation conditions* (esitlustingimused):
  - for f a p.r. operation,

  $$\vdash_T \bar{f}[\bar{m_1}, \ldots, \bar{m_n}] \doteq \bar{m} \text{ iff } f(m_1, \ldots, m_n) = m$$

  - for p a p.r. relation,

  $$\vdash_T \bar{p}[\bar{m_1}, \ldots, \bar{m_n}] \text{ iff } p(m_1, \ldots, m_n)$$

  ($\bar{m}$ denotes the representation of m.)

- **Fact:** The terms and sentences (and schematic terms and schematic sentences) of any rich language L (with denumerable signature) are effectively enumerable by natural numbers so that all important syntatic operactions on them reduce to p.r. operations on numbers (*Gödel numbers*).

- **Consequence:** Because of the representability of natural numbers in L, both the terms and sentences of L therefore translate to L-numerals (*codes*).

  $\ulcorner m \urcorner$ denotes the code of m.

  In powerful L-theories, facts about important operations and relations concerning codes are reflected quite well since the presentation conditions hold.

- **Convention:** From now on, saying "language", we always mean a rich language, and saying "theory", we always mean a powerful theory.

- **Diagonalization Lemma:** Given a language L, one can for any schematic L-sentence P effectively find a sentence S s.t. $\models S \equiv P[\ulcorner S \urcorner]$ and, for any L-theory T, $\vdash_T S \equiv P[\ulcorner S \urcorner]$.

- **Proof:** Instantiating schematic L-sentences with L-numerals is a p.r. operation, reduced to Gödel numbers thus a p.r. operation on numbers, hence representable in L. Let subst be the schematic L-term representing it. Then $\models \mathsf{subst}[\ulcorner Q \urcorner, t] \doteq \ulcorner Q[t] \urcorner$ for any schematic L-sentence Q and any numeral t. For an L-theory T, we have $\vdash_T \mathsf{subst}[\ulcorner Q \urcorner, t] \doteq \ulcorner Q[t] \urcorner$ by the presentation conditions.

  Consider any schematic L-sentence P. Let D be the diagonal schematic L-sentence given by $D[t] := P[\mathsf{subst}[t, t]]$.

  Set $S := D[\ulcorner D \urcorner]$. Then

$$\models S \equiv P[\ulcorner S \urcorner] \text{ and } \vdash_T S \equiv P[\ulcorner S \urcorner]$$

  since by the definitions of S and D, $S \equiv P[\ulcorner S \urcorner]$ is identical to $P[\mathsf{subst}[\ulcorner D \urcorner, \ulcorner D \urcorner]] \equiv P[\ulcorner D[\ulcorner D \urcorner] \urcorner]$.

- **Tarski's theorem about non-representability of truth.** Given a language L, truth of L-sentences is non-representable in L: there is no schematic L-sentence True in L s.t.

$$\models A \text{ iff } \models \mathsf{True}[\ulcorner A \urcorner]$$

- **Proof.** Suppose a schematic L-sentence True with the stated property exists.

  Then, applying the Diagonalization Lemma to the schematic L-sentence $\neg\mathsf{True}$, we can produce an L-sentence Tarski such that $\models \mathsf{Tarski} \equiv \neg\mathsf{True}[\ulcorner \mathsf{Tarski} \urcorner]$ which has the effect that $\models \mathsf{Tarski}$ iff $\not\models \mathsf{True}[\ulcorner \mathsf{Tarski} \urcorner]$, which, by our assumption about True, happens iff $\not\models \mathsf{Tarski}$.

  Hence Tarski is a sentence stating its own falsity, a "liar". Independent of whether Tarski is true <u>or</u> false, it is true <u>and</u> false, which cannot be.

- **Gödel's theorem about representability of theoremhood.** Given a language L, theoremhood in an *axiomatized* L-theory T is effectively representable in L: one can effectively find a schematic sentence Thm_T in L s.t.

$$\vdash_T A \text{ iff } \models \mathsf{Thm}_T[\ulcorner A \urcorner]$$

- **Proof:** For an axiomatized L-theory T, the relation of a sequence of L-sentences being a T-proof of a L-sentence is a p.r. relation, reduced to Gödel numbers, thus a p.r. relation on numbers, thus effectively representable in L. Let Proof_T be the schematic L-sentence representing it.

  Thm_T is constructed by letting $\mathsf{Thm}_T[t] := \exists x.\ \mathsf{Nat}[x] \wedge \mathsf{Proof}_T[x, t]$.

- **Lemma (Gödel):** Given a language L, each <u>axiomatized</u> L-theory T satisfies the following derivability conditions (tuletatavustingimused):

  D1 $\vdash_T A$ implies $\vdash_T \mathsf{Thm}_T[\ulcorner A \urcorner]$ (the theory is positively introspective),

  D2 $\vdash_T \mathsf{Thm}_T[\ulcorner A \supset B \urcorner] \supset (\mathsf{Thm}_T[\ulcorner A \urcorner] \supset \mathsf{Thm}_T[\ulcorner B \urcorner])$ (the theory knows it's closed under modus ponens),

  D3 $\vdash_T \mathsf{Thm}_T[\ulcorner A \urcorner] \supset \mathsf{Thm}_T[\ulcorner \mathsf{Thm}_T[\ulcorner A \urcorner] \urcorner]$ (the theory knows it is positively introspective).

- **Proof:** Hard work (unrewarding).

- **Corollary:** Given a language L, a sound <u>axiomatized</u> L-theory T is necessarily semantically incomplete (and hence because of the assumption of soundness also syntactically incomplete).
- **Proof:** If some L-theory T was both sound and semantically complete, T-theoremhood of L-sentences would be the same as truth. But one is L-representable, the other is not.

- **Gödel's first incompleteness theorem:** Given a language L, for an <u>axiomatized</u> L-theory T, one can effectively find an L-sentence Godel_T s.t.
  - if T is consistent, then $\nvdash_T \mathsf{Godel}_T$, but $\models \mathsf{Godel}_T$ (so T is semantically incomplete),
  - if T is omega-consistent, then $\nvdash_T \neg\mathsf{Godel}_T$ (so T is also syntactically incomplete).
- **Proof:** For an axiomatized L-theory T, we know that a schematic L-sentence Thm_T exists s.t. $\vdash_T A$ iff $\models \mathsf{Thm}_T[\ulcorner A \urcorner]$.

  Using the Diagonalization Lemma, we construct Godel_T as an L-sentence s.t.

  $$\models \mathsf{Godel}_T \equiv \neg\mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner] \text{ and } \vdash_T \mathsf{Godel}_T \equiv \neg\mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner]$$

  (so informally Godel_T says it's a non-T-theorem and that's a T-theorem).

  Assume T is consistent. Suppose $\vdash_T \mathsf{Godel}_T$. Then, by D1, also $\vdash_T \mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner]$. But then, by the construction of Godel_T, $\vdash_T \neg\mathsf{Godel}_T$, which contradicts consistency.

  Suppose $\not\models \mathsf{Godel}_T$, then by the construction of Godel_T, $\models \mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner]$. This is by the construction of Thm_T, equivalent to $\vdash_T \mathsf{Godel}_T$, but we already saw $\vdash_T \neg\mathsf{Godel}_T$, so again we are contradicting consistency.

- **Remark:** Note that while Tarski is an antinomic sentence, it must not exist; Godel_T is merely paradoxical, its existence looks potentially troublesome, but there is nothing harmful about it.

- **Gödel's second incompleteness theorem:** Given a language L, for an <u>axiomatized</u> L-theory T, if T is consistent, than

$$\nvdash_T \mathsf{Cons}_T$$

where $\mathsf{Cons}_T := \neg\mathsf{Thm}_T[\ulcorner\bot\urcorner]$ (which says T is consistent). (So consistency of a consitent axiomatized theory T is not a T-theorem.)

- **Proof:**

Assume T is a consistent axiomatized L-theory. By the construction of Godel_T, we have

$$\vdash_T \mathsf{Godel}_T \supset \neg\mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner]$$

From this, by D1, we get

$$\vdash_T \mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner \supset \neg\mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner]]$$

from where, by D2, we further get

$$\vdash_T \mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner] \supset \mathsf{Thm}_T[\ulcorner\neg\mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner]\urcorner]$$

But by D3 we also have

$$\vdash_T \mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner] \supset \mathsf{Thm}_T[\ulcorner\mathsf{Thm}_T[\ulcorner\mathsf{Godel}_T\urcorner]\urcorner]$$

Combining the last two using D2 and the construction of Cons_T, we get

$$\vdash_T \mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner] \supset \neg \mathsf{Cons}_T$$

which of course gives

$$\vdash_T \mathsf{Cons}_T \supset \neg \mathsf{Thm}_T[\ulcorner \mathsf{Godel}_T \urcorner]$$

Together with the construction of Godel_T again (the second half of the equivalence), this yields

$$\vdash_T \mathsf{Cons}_T \supset \mathsf{Godel}_T$$

If now it were the case that $\vdash_T \mathsf{Cons}_T$, then also $\vdash_T \mathsf{Godel}_T$, but since T is consistent, the First Incompleteness Theorem tell us the that $\nvdash_T \mathsf{Godel}_T$.