RECURSIVELY-DEFINED LOGICAL THEORIES ARE WELL-DEFINED (BRIEF TECHNICAL NOTE)

NISAN STIENNON

This document is part of a collection of quick writeups of results from the December 2013 MIRI research workshop, written during or directly after the workshop. It describes work by Will Sawin, with some contributions from Benja Fallenstein.

Will Sawin and Benja Fallenstein proved this at the December 2013 workshop, and Nisan Stiennon wrote it up. It might already be in the literature.

The "waterfall" in the Tiling Agents paper is a family of theories $\mathrm{PA}+n$ with definitions like

$$\mathrm{PA}+n := \mathrm{PA} \cup \{\Box_{\mathrm{PA}+(n+1)}\phi \to \phi \mid \text{formulas } \phi\} \tag{1}$$

This definition is circular. To make it rigorous, one can show that there exists a unique sequence of theories that agrees with this definition. To show existence, one can use diagonalization to define all the proof predicates and axiom schemas at once. To show uniqueness, we will view (1) as a sort of equation, and then prove that a solution to it must be unique.

Suppose we have a countable indexing set I and an equation of the form

$$T_i \simeq \mathrm{PA} + \cup_k \mathrm{Ax}_i(k) \tag{2}$$

where $\mathrm{Ax}_i(k)$ is the k^{th} extra axiom of T_i. We want these axioms to have terms like $\Box_{T_j}\phi$, so extend the language of PA with a binary predicate symbol $P(j,n)$, where we interpret $P(j, \ulcorner\phi\urcorner)$ as saying that ϕ is provable in T_j, and make $\mathrm{Ax}_i(k)$ be a sentence in this extended langauge. Also, we demand that $\ulcorner\mathrm{Ax}_i(k)\urcorner$ be a recursive function of i and k.

Given a recursively enumerable family of theories $\{U_i\}_{i\in I}$, write $\mathrm{Prv}_U(i,n)$ for the provability predicate asserting that n is the Gödel number of a formula provable in U_i; that is, $\mathrm{Prv}_U(i,\ulcorner\phi\urcorner)$ is equivalent to $\Box_{U_i}\phi$. Say that $\{U_i\}_{i\in I}$ is a *solution* to (2) if

$$\mathrm{PA} \vdash \forall i\big(U_i \simeq \mathrm{PA} \cup_k \mathrm{Ax}_i(k)[\mathrm{Prv}_U/P]\big)$$

where the $\simeq$ symbol means equivalence of theories — $A \simeq B$ stands for $\forall\ulcorner\phi\urcorner.\ \Box_A\phi \leftrightarrow \Box_B\phi$. (More explicitly, the notation $\phi[\mathrm{Prv}_U/P]$ means that we replace all subformulas of the form $P(X,Y)$, where X and Y are terms, by $\mathrm{Prv}_U(X,Y)$.)

Theorem 1. *Suppose $\{U_i\}$ and $\{V_i\}$ are solutions to (2). Then* $\mathrm{PA} \vdash \forall i.\ U_i \simeq V_i$.

Proof. The proof proceeds by Löb's theorem. Let

$$\chi := \Box_{\mathrm{PA}}(\forall i.\ U_i \simeq V_i)$$

Then

$$\begin{aligned}
&\mathrm{PA} + \chi \vdash \Box_{\mathrm{PA}}(\forall i.\ U_i \simeq V_i) \\
&\mathrm{PA} + \chi \vdash \Box_{\mathrm{PA}}(\forall i, \ulcorner\phi\urcorner.\ \Box_{U_i}\phi \leftrightarrow \Box_{V_i}\phi) \\
&\mathrm{PA} + \chi \vdash \forall j.\ \Box_{U_j}(\forall i, \ulcorner\phi\urcorner.\ \Box_{U_i}\phi \leftrightarrow \Box_{V_i}\phi) \\
&\mathrm{PA} + \chi \vdash \forall j, k.\ \Box_{U_j}(\mathrm{Ax}_j(k)[\mathrm{Prv}_U\,/P] \leftrightarrow \mathrm{Ax}_j(k)[\mathrm{Prv}_U\,/P]) \\
&\mathrm{PA} + \chi \vdash \forall j, k.\ \Box_{U_j}(\mathrm{Ax}_j(k)[\mathrm{Prv}_V\,/P]) \\
&\mathrm{PA} + \chi \vdash \forall j, \ulcorner\phi\urcorner.\ \Box_{U_j}\phi \rightarrow \Box_{V_j}\phi
\end{aligned}$$

By symmetry, we also have

$$\begin{aligned}
&\mathrm{PA} + \chi \vdash \forall j, \ulcorner\phi\urcorner.\ \Box_{U_j}\phi \leftrightarrow \Box_{V_j}\phi \\
&\mathrm{PA} + \chi \vdash \forall j.\ (U_j \simeq V_j)
\end{aligned}$$

So by Löb's theorem, we conclude that PA $\vdash \forall i.\ U_i \simeq V_i$. (And since PA is sound, the solutions $\{U_i\}$ and $\{V_i\}$ are in fact equivalent.)

□

Of course, if $Ax_i(k)$ uses $P(j, n)$ only for $j < i$ in some well-founded ordering of I, then this result follows from induction on that ordering. What's interesting is that it works just as well if the recursion is not well-founded.

The shadiest step of that proof is where we go from

$$\Box_{U_j}(\forall i, \ulcorner\phi\urcorner.\ \Box_{U_i}\ulcorner\phi\urcorner \leftrightarrow \Box_{V_i}\ulcorner\phi\urcorner)$$

to

$$\Box_{U_j}(\mathrm{Ax}_j(k)[\mathrm{Prv}_U\,/P] \leftrightarrow \mathrm{Ax}_j(k)[\mathrm{Prv}_V\,/P])$$

For that we need the following lemma, which is provable in PA:

Lemma 1. *If T is a theory in first-order logic, $P(x, y)$ is a predicate symbol, and $\psi_1(x, y)$ and $\psi_2(x, y)$ are formulas such that $T \vdash \forall x, y.\ \psi_1(x, y) \leftrightarrow \psi_2(x, y)$, then for all formulas ϕ, $T \vdash \phi[\psi_1/P] \leftrightarrow \phi[\psi_2/P]$.*

The theorem has a number of immediate corollaries:

Corollary 1. *Any system of theories satisfying the "naive waterfall" recurrence*

$$T_n \simeq \mathrm{PA} + \cup_\phi\, \Box_{T_{n+1}}\phi \rightarrow \phi$$

is inconsistent.

Proof. $U_n := \{\bot\}$ is a solution. Therefore it is the unique solution.

(Perhaps more illuminatingly, you could use diagonalization to construct the theory $U := \mathrm{PA} \cup_\phi \Box_U\phi \rightarrow \phi$. Then $U_n = U$ is a solution and by Gödel's theorem, this solution is inconsistent.) □

Amusingly, we can also prove Gödel's second incompleteness theorem and Löb's theorem:

Corollary 2. *Any extension of arithmetic that satisfies an equation of the form*

$$T \simeq T + \mathrm{Con}(T)$$

is inconsistent.

Proof. If there is such a theory U, we can rewrite the equation as

$$T \simeq \mathrm{PA} + \cdots + \mathrm{Con}(T)$$

where the ellipsis represents an axiom schema for U. Setting $V := \{\perp\}$ gives a solution, and therefore the unique solution. □

Corollary 3. *Any extension of arithmetic satisfying $T \vdash \Box_T\phi \to \phi$, for some ϕ, satisfies $T \vdash \phi$.*

Proof. Any such U must be a solution to

$$T \simeq \mathrm{PA} + \cdots + \Box_T\phi \to \phi$$

where the ellipsis stands for the axioms of U. Setting $V := U + \phi$ gives another solution. Therefore $U \simeq U + \phi$, and so $U \vdash \phi$. □

Thus one might view this theorem as an alternative formulation of Löb's Theorem.