Notes on the incompleteness theorems

Makoto Kikuchi and Taishi Kurahashi

Kobe University, Japan

Logic Colloquium
July 11-16, 2011
Barcelona

We present three theorems relating to the following topics about the incompleteness theorems.

1. **The self-referentiality of the independent propositions.**
2. **The formalization of the proofs of the first incompleteness theorem.**
3. **The definability of the truth in the models of arithmetic.**

In this talk, we fix the following objects.

- T: an r.e. extension of PA in the language of arithmetic (denote $\mathcal{L}_A$);
- $\mathrm{Proof}_T(x, y)$: a canonical Δ_1 formula which states that "y is a proof of a formula x in T".

$\mathrm{Pr}_T(x) :\equiv \exists y \mathrm{Proof}_T(x, y)$.
$\mathrm{Con}(T) :\equiv \neg \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner)$.

We present three theorems relating to the following topics about the incompleteness theorems.

1. **The self-referentiality of the independent propositions.**
2. **The formalization of the proofs of the first incompleteness theorem.**
3. **The definability of the truth in the models of arithmetic.**

In this talk, we fix the following objects.

- T: an r.e. extension of PA in the language of arithmetic (denote $\mathcal{L}_A$);
- $\mathrm{Proof}_T(x, y)$: a canonical Δ_1 formula which states that "y is a proof of a formula x in T".

$\mathrm{Pr}_T(x) :\equiv \exists y \mathrm{Proof}_T(x, y)$.
$\mathsf{Con}(T) :\equiv \neg \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner)$.

The self-referentiality of the independent propositions.

Yablo's paradox

- **Let $Y_0, Y_1, \ldots,$ be an infinite sequence of propositions.**
- **Each Y_i states that "For every $j > i$, Y_j is false".**
- **Then we cannot determine whether Y_i is true or false.**

Yablo, 1993

Yablo's paradox is not self-referential.

Priest, 1997

- **Yablo's paradox is self-referential.**
- **The first incompleteness theorem is proved by formalizing Yablo's paradox.**

The self-referentiality of the independent propositions.

Yablo's paradox

- **Let $Y_0, Y_1, \ldots,$ be an infinite sequence of propositions.**
- **Each Y_i states that "For every $j > i$, Y_j is false".**
- **Then we cannot determine whether Y_i is true or false.**

Yablo, 1993

Yablo's paradox is not self-referential.

Priest, 1997

- **Yablo's paradox is self-referential.**
- **The first incompleteness theorem is proved by formalizing Yablo's paradox.**

The self-referentiality of the independent propositions.

To formalize Yablo's paradox, we need the following version of the diagonalization lemma.

The diagonalization lemma

$\forall\varphi(x,y,z)$: $\mathcal{L}_A$-formula whose only the free variables are x, y, z,
$\exists\psi(x)$: $\mathcal{L}_A$-formula with only the free variable x s.t.

$$T \vdash \forall x(\psi(x) \leftrightarrow \forall y\varphi(x,y,\ulcorner\psi(\dot{y})\urcorner)).$$

Here $\ulcorner\psi(\dot{x})\urcorner$ is a numeral of the Gödel number of the sentence which is obtained by substituting x to the formula $\psi(v)$.

A formalization of Yablo's paradox

Let $Y(x)$ be an $\mathcal{L}_A$-formula which satisfies the following equivalence:

$$T \vdash \forall x(Y(x) \leftrightarrow \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y})\urcorner)).$$

The self-referentiality of the independent propositions.

To formalize Yablo's paradox, we need the following version of the diagonalization lemma.

The diagonalization lemma

$\forall\varphi(x,y,z)$: $\mathcal{L}_A$-formula whose only the free variables are x,y,z,
$\exists\psi(x)$: $\mathcal{L}_A$-formula with only the free variable x s.t.

$$T \vdash \forall x(\psi(x) \leftrightarrow \forall y \varphi(x,y,\ulcorner\psi(\dot{y})\urcorner)).$$

Here $\ulcorner\psi(\dot{x})\urcorner$ is a numeral of the Gödel number of the sentence which is obtained by substituting x to the formula $\psi(v)$.

A formalization of Yablo's paradox

Let $Y(x)$ be an $\mathcal{L}_A$-formula which satisfies the following equivalence:

$$T \vdash \forall x(Y(x) \leftrightarrow \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y})\urcorner)).$$

The self-referentiality of the independent propositions.

To formalize Yablo's paradox, we need the following version of the diagonalization lemma.

The diagonalization lemma

$\forall\varphi(x,y,z)$: $\mathcal{L}_A$-formula whose only the free variables are x,y,z,
$\exists\psi(x)$: $\mathcal{L}_A$-formula with only the free variable x s.t.

$$T \vdash \forall x(\psi(x) \leftrightarrow \forall y \varphi(x,y,\ulcorner\psi(\dot{y})\urcorner)).$$

Here $\ulcorner\psi(\dot{x})\urcorner$ is a numeral of the Gödel number of the sentence which is obtained by substituting x to the formula $\psi(v)$.

A formalization of Yablo's paradox

Let $Y(x)$ be an $\mathcal{L}_A$-formula which satisfies the following equivalence:

$$T \vdash \forall x(Y(x) \leftrightarrow \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y})\urcorner)).$$

The self-referentiality of the independent propositions.

The first incompleteness theorem (Priest, 1997)

For any $n \in \mathbb{N}$,

1. **If T is consistent, then $T \nvdash Y(\bar{n})$.**
2. **If T is Σ_1-sound, then $T \nvdash \neg Y(\bar{n})$.**

We proved the second incompleteness theorem by this formalization of Yablo's paradox.

The second incompleteness theorem (M.K. and T.K.)

$$T \vdash \mathrm{Con}(T) \to \forall x Y(x),$$

thus if T is consistent, then $T \nvdash \mathrm{Con}(T)$.

The self-referentiality of the independent propositions.

The first incompleteness theorem (Priest, 1997)

For any $n \in \mathbb{N}$,

1. **If T is consistent, then $T \nvdash Y(\bar{n})$.**
2. **If T is Σ_1-sound, then $T \nvdash \neg Y(\bar{n})$.**

We proved the second incompleteness theorem by this formalization of Yablo's paradox.

The second incompleteness theorem (M.K. and T.K.)

$$T \vdash \mathrm{Con}(T) \to \forall x Y(x),$$

thus if T is consistent, then $T \nvdash \mathrm{Con}(T)$.

The self-referentiality of the independent propositions.

The first incompleteness theorem (Priest, 1997)

For any $n \in \mathbb{N}$,

1. **If T is consistent, then $T \nvdash Y(\bar{n})$.**
2. **If T is Σ_1-sound, then $T \nvdash \neg Y(\bar{n})$.**

We proved the second incompleteness theorem by this formalization of Yablo's paradox.

The second incompleteness theorem (M.K. and T.K.)

$$T \vdash \mathsf{Con}(T) \to \forall x Y(x),$$

thus if T is consistent, then $T \nvdash \mathsf{Con}(T)$.

The self-referentiality of the independent propositions.

An outline of a proof.

Formalize the proof of the first incompleteness theorem using Yablo's paradox, that is,

$$T \vdash \mathsf{Con}(T) \to \forall x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{x}) \urcorner).$$

Then

$$\begin{aligned} T &\vdash \mathsf{Con}(T) \to \forall x \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y}) \urcorner), \\ T &\vdash \mathsf{Con}(T) \to \forall x Y(x). \end{aligned}$$

- **The problem whether our proofs of the incompleteness theorems formalizing Yablo's paradox are self-referential is as difficult as the problem of the self-referentiality of Yablo's paradox.**

The self-referentiality of the independent propositions.

An outline of a proof.

Formalize the proof of the first incompleteness theorem using Yablo's paradox, that is,

$$T \vdash \mathsf{Con}(T) \to \forall x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{x}) \urcorner).$$

Then

$$\begin{aligned} T &\vdash \mathsf{Con}(T) \to \forall x \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y}) \urcorner), \\ T &\vdash \mathsf{Con}(T) \to \forall x Y(x). \end{aligned}$$

- **The problem whether our proofs of the incompleteness theorems formalizing Yablo's paradox are self-referential is as difficult as the problem of the self-referentiality of Yablo's paradox.**

The self-referentiality of the independent propositions.

An outline of a proof.

Formalize the proof of the first incompleteness theorem using Yablo's paradox, that is,

$$T \vdash \mathsf{Con}(T) \to \forall x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{x}) \urcorner).$$

Then

$$\begin{aligned} T &\vdash \mathsf{Con}(T) \to \forall x \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y}) \urcorner), \\ T &\vdash \mathsf{Con}(T) \to \forall x Y(x). \end{aligned}$$

- The problem whether our proofs of the incompleteness theorems formalizing Yablo's paradox are self-referential is as difficult as the problem of the self-referentiality of Yablo's paradox.

The self-referentiality of the independent propositions.

An outline of a proof.

Formalize the proof of the first incompleteness theorem using Yablo's paradox, that is,

$$T \vdash \mathsf{Con}(T) \to \forall x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{x}) \urcorner).$$

Then

$$\begin{aligned} T &\vdash \mathsf{Con}(T) \to \forall x \forall y > x \neg \mathrm{Pr}_T(\ulcorner Y(\dot{y}) \urcorner), \\ T &\vdash \mathsf{Con}(T) \to \forall x Y(x). \end{aligned}$$

- **The problem whether our proofs of the incompleteness theorems formalizing Yablo's paradox are self-referential is as difficult as the problem of the self-referentiality of Yablo's paradox.**

The formalization of the proofs of the first incompleteness theorem.

The proof of Gödel's second incompleteness theorem is curried out by proving the formalization of the first incompleteness theorem.

φ **: Gödel sentence,**

$$T \vdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner).$$

We proved the existence of an independent Π_1 sentence whose unprovability is not formalizable.

Theorem (M.K. and T.K.)

The following are equivalent:

- $T \nvdash \neg \mathsf{Con}(T)$;
- $\exists \varphi$: **independent Π_1 sentence s.t.** $T \nvdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The formalization of the proofs of the first incompleteness theorem.

The proof of Gödel's second incompleteness theorem is curried out by proving the formalization of the first incompleteness theorem.

φ **: Gödel sentence,**

$$T \vdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner).$$

We proved the existence of an independent Π_1 sentence whose unprovability is not formalizable.

Theorem (M.K. and T.K.)

The following are equivalent:

- $T \nvdash \neg \mathsf{Con}(T)$;
- $\exists \varphi$: **independent Π_1 sentence s.t.** $T \nvdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The formalization of the proofs of the first incompleteness theorem.

The proof of Gödel's second incompleteness theorem is curried out by proving the formalization of the first incompleteness theorem.

φ : Gödel sentence,

$$T \vdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner).$$

We proved the existence of an independent Π_1 sentence whose unprovability is not formalizable.

Theorem (M.K. and T.K.)

The following are equivalent:

- $T \nvdash \neg \mathsf{Con}(T)$;
- $\exists \varphi$: independent Π_1 sentence s.t. $T \nvdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The formalization of the proofs of the first incompleteness theorem.

An outline of a proof.

Assume that $T \nvdash \neg \mathsf{Con}(T)$.

Let φ be an $\mathcal{L}_A$-sentence obtained by the following equivalence:

$$T \vdash \varphi \quad \leftrightarrow \quad \forall x(\mathrm{Proof}_T(\ulcorner \varphi \urcorner, x) \to \exists y \leq x \mathrm{Proof}_T(\ulcorner \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner, y)).$$

Then

- φ **is equivalent to a Π_1 sentence;**
- φ **is independent from T;**
- $T \nvdash \mathsf{Con}(T) \to \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The formalization of the proofs of the first incompleteness theorem.

An outline of a proof.

Assume that $T \nvdash \neg \mathsf{Con}(T)$.
Let φ be an $\mathcal{L}_A$-sentence obtained by the following equivalence:

$$T \vdash \varphi \quad \leftrightarrow \quad \forall x(\mathrm{Proof}_T(\ulcorner \varphi \urcorner, x) \rightarrow \exists y \leq x \mathrm{Proof}_T(\ulcorner \mathsf{Con}(T) \rightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner, y)).$$

Then

- φ **is equivalent to a Π_1 sentence;**
- φ **is independent from T;**
- $T \nvdash \mathsf{Con}(T) \rightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The formalization of the proofs of the first incompleteness theorem.

Immediately, we have the following Corollary.

Corollary

$\exists \mathcal{M}$: model of T, $\exists \varphi$: Π_1 sentence s.t.

- φ is independent from T,
- $\mathcal{M} \models \mathsf{Con}(T)$,
- $\mathcal{M} \models \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$.

The definability of the truth in the models of arithmetic.

Theorem (Tarski, 1936)

There is no $\mathcal{L}_A$-formula $\Phi(x)$ s.t.
$\forall\varphi$: $\mathcal{L}_A$-formula,

$$\mathbb{N} \models \varphi \Leftrightarrow \mathbb{N} \models \Phi(\ulcorner\varphi\urcorner).$$

Tarski's theorem can be easily extended to any model of T.

An extended Tarski's Theorem

$\forall\mathcal{M}$: model of T,
there is no $\mathcal{L}_A$-formula $\Phi(x)$ s.t.
$\forall\varphi$: $\mathcal{L}_A$-formula,

$$\mathcal{M} \models \varphi \Leftrightarrow \mathcal{M} \models \Phi(\ulcorner\varphi\urcorner).$$

The definability of the truth in the models of arithmetic.

Theorem (Tarski, 1936)

There is no $\mathcal{L}_A$-formula $\Phi(x)$ s.t.
$\forall\varphi$: $\mathcal{L}_A$-formula,

$$\mathbb{N} \models \varphi \Leftrightarrow \mathbb{N} \models \Phi(\ulcorner\varphi\urcorner).$$

Tarski's theorem can be easily extended to any model of T.

An extended Tarski's Theorem

$\forall\mathcal{M}$: model of T,
there is no $\mathcal{L}_A$-formula $\Phi(x)$ s.t.
$\forall\varphi$: $\mathcal{L}_A$-formula,

$$\mathcal{M} \models \varphi \Leftrightarrow \mathcal{M} \models \Phi(\ulcorner\varphi\urcorner).$$

The definability of the truth in the models of arithmetic.

We proved that for any model $\mathcal{M}$ of T, the truth of $\mathcal{M}$ is defined by an $\mathcal{L}_A$-formula in another certain model of T.

Theorem (M.K. and T.K.)

$\forall \mathcal{M}$: model of T,
$\exists \mathcal{N}$: model of T $\exists \Psi(x)$: $\mathcal{L}_A$-formula s.t.
$\forall \varphi$: $\mathcal{L}_A$-formula,

$$\mathcal{M} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner \varphi \urcorner).$$

Corollary

$\exists \mathcal{N}$: model of T $\exists \Psi(x)$: $\mathcal{L}_A$-formula s.t.
$\forall \varphi$: $\mathcal{L}_A$-formula,

$$\mathbb{N} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner \varphi \urcorner).$$

The definability of the truth in the models of arithmetic.

We proved that for any model $\mathcal{M}$ of T, the truth of $\mathcal{M}$ is defined by an $\mathcal{L}_A$-formula in another certain model of T.

Theorem (M.K. and T.K.)

$\forall\mathcal{M}$: **model of** T,
$\exists\mathcal{N}$: **model of** T $\exists\Psi(x)$: $\mathcal{L}_A$**-formula s.t.**
$\forall\varphi$: $\mathcal{L}_A$**-formula,**

$$\mathcal{M} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner\varphi\urcorner).$$

Corollary

$\exists\mathcal{N}$: **model of** T $\exists\Psi(x)$: $\mathcal{L}_A$**-formula s.t.**
$\forall\varphi$: $\mathcal{L}_A$**-formula,**

$$\mathbb{N} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner\varphi\urcorner).$$

The definability of the truth in the models of arithmetic.

We proved that for any model $\mathcal{M}$ of T, the truth of $\mathcal{M}$ is defined by an $\mathcal{L}_A$-formula in another certain model of T.

Theorem (M.K. and T.K.)

$\forall \mathcal{M}$: **model of** T,
$\exists \mathcal{N}$: **model of** T $\exists \Psi(x)$: $\mathcal{L}_A$**-formula s.t.**
$\forall \varphi$: $\mathcal{L}_A$**-formula,**

$$\mathcal{M} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner \varphi \urcorner).$$

Corollary

$\exists \mathcal{N}$: **model of** T $\exists \Psi(x)$: $\mathcal{L}_A$**-formula s.t.**
$\forall \varphi$: $\mathcal{L}_A$**-formula,**

$$\mathbb{N} \models \varphi \Leftrightarrow \mathcal{N} \models \Psi(\ulcorner \varphi \urcorner).$$

The definability of the truth in the models of arithmetic.

An outline of a proof.

- **Define $\varphi^0 :\equiv \neg\varphi$, $\varphi^1 :\equiv \varphi$.**
- $\exists\Psi(x)$: an $\mathcal{L}_A$-formula s.t. $\forall f : \mathbb{N} \to \{0,1\}$, $\{\Psi(\bar{n})^{f(n)} \; : \; n \in \mathbb{N}\}$ is consistent.
- Let $f^* : \mathbb{N} \to \{0,1\}$ be a function defined by

$$\begin{aligned} & f^*(n) = 1 \\ :\Leftrightarrow \quad & n = \ulcorner\varphi\urcorner \text{ for some } \mathcal{L}_A\text{-formula } \varphi \text{ s.t. } \mathcal{M} \models \varphi. \end{aligned}$$

- Then $\{\Psi(\bar{n})^{f^*(n)} \; : \; n \in \mathbb{N}\}$ has a model $\mathcal{N}$.

The definability of the truth in the models of arithmetic.

An outline of a proof.

- **Define** $\varphi^0 :\equiv \neg\varphi$, $\varphi^1 :\equiv \varphi$.
- $\exists \Psi(x)$**: an** $\mathcal{L}_A$**-formula s.t.**
  $\forall f : \mathbb{N} \to \{0,1\}$, $\{\Psi(\bar{n})^{f(n)} \ : \ n \in \mathbb{N}\}$ **is consistent.**
- Let $f^* : \mathbb{N} \to \{0,1\}$ be a function defined by

$$f^*(n) = 1$$
$$:\Leftrightarrow \quad n = \ulcorner \varphi \urcorner \text{ for some } \mathcal{L}_A\text{-formula } \varphi \text{ s.t. } \mathcal{M} \models \varphi.$$

- Then $\{\Psi(\bar{n})^{f^*(n)} \ : \ n \in \mathbb{N}\}$ has a model $\mathcal{N}$.

The definability of the truth in the models of arithmetic.

An outline of a proof.

- Define $\varphi^0 :\equiv \neg\varphi$, $\varphi^1 :\equiv \varphi$.
- $\exists\Psi(x)$: an $\mathcal{L}_A$-formula s.t.
  $\forall f : \mathbb{N} \to \{0,1\}$, $\{\Psi(\bar{n})^{f(n)} \ : \ n \in \mathbb{N}\}$ is consistent.
- Let $f^* : \mathbb{N} \to \{0,1\}$ be a function defined by

$$\begin{aligned} & f^*(n) = 1 \\ :\Leftrightarrow \quad & n = \ulcorner\varphi\urcorner \text{ for some } \mathcal{L}_A\text{-formula } \varphi \text{ s.t. } \mathcal{M} \models \varphi. \end{aligned}$$

- Then $\{\Psi(\bar{n})^{f^*(n)} \ : \ n \in \mathbb{N}\}$ has a model $\mathcal{N}$.

The definability of the truth in the models of arithmetic.

An outline of a proof.

- **Define** $\varphi^0 :\equiv \neg\varphi$, $\varphi^1 :\equiv \varphi$.
- $\exists\Psi(x)$: **an** $\mathcal{L}_A$**-formula s.t.**
  $\forall f : \mathbb{N} \to \{0,1\}$, $\{\Psi(\bar{n})^{f(n)} \ : \ n \in \mathbb{N}\}$ **is consistent.**
- **Let** $f^* : \mathbb{N} \to \{0,1\}$ **be a function defined by**

$$
\begin{aligned}
& f^*(n) = 1 \\
:\Leftrightarrow \quad & n = \ulcorner\varphi\urcorner \text{ for some } \mathcal{L}_A\text{-formula } \varphi \text{ s.t. } \mathcal{M} \models \varphi.
\end{aligned}
$$

- **Then** $\{\Psi(\bar{n})^{f^*(n)} \ : \ n \in \mathbb{N}\}$ **has a model** $\mathcal{N}$**.**

References

- **G. Boolos,** *The logic of provability*. **Cambridge University Press, Cambridge, 1993.**
- **P. Lindström,** *Aspects of Incompleteness*. **Lecture Notes in Logic 10. Springer, 1997.**
- **G. Priest,** *Yablo's paradox*. **Analysis, vol.57 (1997), pp.236–242.**
- **R. Solovay,** *Provability interpretations of modal logic*. **Israel J. Math. 25 (1976), no. 3-4. pp.287–304.**
- **S. Yablo,** *Paradox without self-reference*. **Analysis, vol.53 (1993), pp.251–252.**