Rosser sentences and Rosser logics

Vítězslav Švejdar

Department of Logic

Prague Gathering of Logicians
January 26–27, 2017

Outline

Gödel's sentence, Rosser's sentence

Other prominent self-referential constructions

Provability logic GL and its applications

Rosser logics

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\mathrm{Pr}(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$.

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\mathrm{Pr}(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$. Then:

$$T \vdash \rho \equiv \exists y (\mathrm{Proof}(\overline{\neg\rho}, y) \; \& \; \forall v < y \, \neg \mathrm{Proof}(\overline{\rho}, v))$$

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\mathrm{Pr}(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$. Then:

$$T \vdash \rho \equiv \exists y(\mathrm{Proof}(\overline{\neg\rho}, y) \ \& \ \forall v{<}y \neg\mathrm{Proof}(\overline{\rho}, v))$$

$$T \vdash \rho \equiv \exists y \mathrm{Proof}(\overline{\neg\rho}, y) \preceq \exists y \mathrm{Proof}(\overline{\rho}, y)$$

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \Pr(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\Pr(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \Pr(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\Pr(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$. Then:

$$T \vdash \rho \equiv \exists y(\mathrm{Proof}(\overline{\neg\rho}, y) \;\&\; \forall v{<}y \neg\mathrm{Proof}(\overline{\rho}, v))$$

$$T \vdash \rho \equiv \exists y \mathrm{Proof}(\overline{\neg\rho}, y) \preceq \exists y \mathrm{Proof}(\overline{\rho}, y)$$

$$T \vdash \rho \equiv \Pr(\overline{\neg\rho}) \preceq \Pr(\overline{\rho})$$

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\mathrm{Pr}(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$. Then:

$$T \vdash \rho \equiv \exists y(\mathrm{Proof}(\overline{\neg\rho}, y) \;\&\; \forall v{<}y \neg\mathrm{Proof}(\overline{\rho}, v))$$

$$T \vdash \rho \equiv \exists y \mathrm{Proof}(\overline{\neg\rho}, y) \preceq \exists y \mathrm{Proof}(\overline{\rho}, y)$$

$$T \vdash \rho \equiv \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$$

Are different ways to write the sentence.

Gödel's sentence, Rosser's sentence

Gödel's sentence: $T \vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$.

Ingredients of independence proofs

(i) The provability predicate $\mathrm{Pr}(x)$ of a theory T *defines* the set $\mathrm{Thm}(T)$ of all provable sentences: $T \vdash \varphi \Leftrightarrow \mathbb{N} \models \mathrm{Pr}(\overline{\varphi})$.
(ii) Σ-completeness: if σ is a Σ_1-sentence valid in $\mathbb{N}$, then $T \vdash \sigma$.

Rosser's sentence

The provability predicate $\mathrm{Pr}(x)$ has the form $\exists y \mathrm{Proof}(x, y)$, where $\exists y \mathrm{Proof}(x, y) \in \Delta_1$ is a *proof predicate*, which defines the proof relation: m is a proof of φ iff $\mathbb{N} \models \mathrm{Proof}(\overline{\varphi}, \overline{m})$. Then:

$$T \vdash \rho \equiv \exists y(\mathrm{Proof}(\overline{\neg\rho}, y) \;\&\; \forall v{<}y \neg\mathrm{Proof}(\overline{\rho}, v))$$

$$T \vdash \rho \equiv \exists y \mathrm{Proof}(\overline{\neg\rho}, y) \preceq \exists y \mathrm{Proof}(\overline{\rho}, y)$$

$$T \vdash \rho \equiv \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}),$$ (witness comparison symbols $\preceq, \prec$)

Are different ways to write the sentence.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and
$T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$

or

$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and
$T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$	or	$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.
In the right case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ and $T \vdash \rho$. Inside T, $\mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ yields $\neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which is $\neg\rho$.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$

or

$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.
In the right case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ and $T \vdash \rho$. Inside T, $\mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ yields $\neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which is $\neg\rho$.

A variant proof

If $T \vdash \rho$, then $T \vdash \mathrm{Proof}(\overline{\rho}, \overline{m})$ for some m, and
$T \vdash \neg\mathrm{Proof}(\overline{\neg\rho}, \overline{n})$ for every n.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$

or

$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.
In the right case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ and $T \vdash \rho$. Inside T, $\mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ yields $\neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which is $\neg\rho$.

A variant proof

If $T \vdash \rho$, then $T \vdash \mathrm{Proof}(\overline{\rho}, \overline{m})$ for some m, and $T \vdash \neg\mathrm{Proof}(\overline{\neg\rho}, \overline{n})$ for every n. Then $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$,

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$

or

$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.
In the right case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ and $T \vdash \rho$. Inside T, $\mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ yields $\neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which is $\neg\rho$.

A variant proof

If $T \vdash \rho$, then $T \vdash \mathrm{Proof}(\overline{\rho}, \overline{m})$ for some m, and $T \vdash \neg\mathrm{Proof}(\overline{\neg\rho}, \overline{n})$ for every n. Then $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$, hence $T \vdash \neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which means $T \vdash \neg\rho$.

Independence of Rosser's sentence

A canonical proof

Let $T \vdash \rho$ or $T \vdash \neg\rho$. Then $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ or $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$. Thus

$\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$ and $T \vdash \mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho})$

or

$\mathbb{N} \models \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ and $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$.

In the left case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\neg\rho})$, $T \vdash \neg\rho$ and $T \vdash \rho$.
In the right case we have $\mathbb{N} \models \mathrm{Pr}(\overline{\rho})$ and $T \vdash \rho$. Inside T, $\mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$ yields $\neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which is $\neg\rho$.

A variant proof

If $T \vdash \rho$, then $T \vdash \mathrm{Proof}(\overline{\rho}, \overline{m})$ for some m, and $T \vdash \neg\mathrm{Proof}(\overline{\neg\rho}, \overline{n})$ for every n. Then $T \vdash \mathrm{Pr}(\overline{\rho}) \prec \mathrm{Pr}(\overline{\neg\rho})$, hence $T \vdash \neg(\mathrm{Pr}(\overline{\neg\rho}) \preceq \mathrm{Pr}(\overline{\rho}))$, which means $T \vdash \neg\rho$.
Similarly, if $T \vdash \neg\rho$, then $T \vdash \mathrm{Proof}(\overline{\neg\rho}, \overline{m})$ for some m, and $T \vdash \neg\mathrm{Proof}(\overline{\rho}, \overline{n})$ for every n. Then $T \vdash \rho$.

Aspects and remarks

1. Comparison of Gödel's and Rosser's sentences: some parts of the above reasoning are, but some are not formalizable in the theory itself:
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\gamma})$,
   $T \nvdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\gamma})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\rho})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\rho})$.

Aspects and remarks

1. Comparison of Gödel's and Rosser's sentences: some parts of the above reasoning are, but some are not formalizable in the theory itself:
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\gamma})$,
   $T \nvdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\gamma})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\rho})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\rho})$.
2. Comparison of the canonical and the variant proof: the variant proof is less demanding on resources: Σ-completeness is applied to simpler sentences $(\mathrm{Proof}(\ldots, \overline{n})$ or $\neg\mathrm{Proof}(\ldots, \overline{n}))$, but not to $\mathrm{Pr}(\ldots)$); some induction is involved in the canonical proof.

Aspects and remarks

1. Comparison of Gödel's and Rosser's sentences: some parts of the above reasoning are, but some are not formalizable in the theory itself:
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\gamma})$,
   $T \nvdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\gamma})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\neg\rho})$,
   $T \vdash \mathrm{Con}(T) \rightarrow \neg\mathrm{Pr}(\overline{\rho})$.
2. Comparison of the canonical and the variant proof: the variant proof is less demanding on resources: Σ-completeness is applied to simpler sentences ($\mathrm{Proof}(\ldots, \overline{n})$ or $\neg\mathrm{Proof}(\ldots, \overline{n})$, but not to $\mathrm{Pr}(\ldots)$); some induction is involved in the canonical proof.
3. Uniqueness: the self-reference guarantees the existence of certain sentence, but does not say that it is unique.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \to \kappa$.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \rightarrow \kappa$.

Hájek's sentence

Since GB is finitely axiomatizable, $\{ \varphi \, ; \, \mathrm{GB} \rhd \mathrm{GB} + \varphi \}$ is *RE*.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \rightarrow \kappa$.

Hájek's sentence

Since GB is finitely axiomatizable, $\{ \varphi \,;\, \mathsf{GB} \rhd \mathsf{GB} + \varphi \}$ is *RE*.
Thus one can have χ s.t. $T \vdash \chi \equiv \mathrm{Intp}_{\mathsf{GB}}(\overline{\neg\chi}) \preceq \mathrm{Intp}_{\mathsf{GB}}(\overline{\chi})$.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \rightarrow \kappa$.

Hájek's sentence

Since GB is finitely axiomatizable, $\{\, \varphi\, ;\ \mathrm{GB} \rhd \mathrm{GB} + \varphi \,\}$ is *RE*.
Thus one can have χ s.t. $T \vdash \chi \equiv \mathrm{Intp}_{\mathrm{GB}}(\overline{\neg\chi}) \preceq \mathrm{Intp}_{\mathrm{GB}}(\overline{\chi})$.
Then $\mathrm{GB} \ntriangleright \mathrm{GB} + \chi$ and $\mathrm{GB} \ntriangleright \mathrm{GB} + \neg\chi$.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \rightarrow \kappa$.

Hájek's sentence

Since GB is finitely axiomatizable, $\{\, \varphi \,;\, \mathrm{GB} \rhd \mathrm{GB} + \varphi \,\}$ is *RE*.
Thus one can have χ s.t. $T \vdash \chi \equiv \mathrm{Intp}_{\mathrm{GB}}(\overline{\neg\chi}) \preceq \mathrm{Intp}_{\mathrm{GB}}(\overline{\chi})$.
Then $\mathrm{GB} \not\rhd \mathrm{GB} + \chi$ and $\mathrm{GB} \not\rhd \mathrm{GB} + \neg\chi$.

Mostowski's flexible formula

is a formula $\varphi(x)$ such that $T \cup \{\, \pm\varphi(\overline{0}), \pm\varphi(\overline{1}), \pm\varphi(\overline{2}), \ldots \}$ is consistent for every choice of pluses and minuses.

Some other self-referential sentences

Henkin's sentence: $T \vdash \kappa \equiv \mathrm{Pr}(\kappa)$.

Löb's sentence: $T \vdash \lambda \equiv \mathrm{Pr}(\lambda) \rightarrow \kappa$.

Hájek's sentence

Since GB is finitely axiomatizable, $\{\, \varphi \,;\ \mathrm{GB} \rhd \mathrm{GB} + \varphi \,\}$ is *RE*. Thus one can have χ s.t. $T \vdash \chi \equiv \mathrm{Intp}_{\mathrm{GB}}(\overline{\neg\chi}) \preceq \mathrm{Intp}_{\mathrm{GB}}(\overline{\chi})$. Then $\mathrm{GB} \not\rhd \mathrm{GB} + \chi$ and $\mathrm{GB} \not\rhd \mathrm{GB} + \neg\chi$.

Mostowski's flexible formula

is a formula $\varphi(x)$ such that $T \cup \{\, \pm\varphi(\overline{0}), \pm\varphi(\overline{1}), \pm\varphi(\overline{2}), \ldots \,\}$ is consistent for every choice of pluses and minuses.

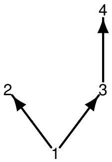
Hájková-Hájek

$\mathrm{PA} \vdash \mu \equiv \forall y(\mathrm{Con}(\mathrm{PA}{\restriction} y + \overline{\mu}) \rightarrow \neg\alpha(\overline{\mu}, y))$, where $\mathrm{PA}{\restriction} y$ are axioms of PA less than y, and $\alpha(x, y) \in \Delta_0$ defines a set A of consistent sentences. Then $\mu \notin A$, but $\mathrm{PA} \rhd \mathrm{PA} + \mu$.

Self-referential sentences (continued)

Embedding a Kripke model to PA (Solovay)

Let $k = \langle W, R \rangle$, $W = \{1, .., n\}$. Put $S(i) = \{\, j \,;\, i\, R\, j \,\}$. Example:



$S(1) = \{2, 3, 4\}$, $S(3) = \{4\}$, $S(2) = S(4) = \emptyset$.

Then consistent sentences $\lambda_1, .., \lambda_n$ such that

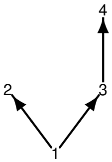
$$\text{PA} \vdash \lambda_i \;\rightarrow\; \bigwedge_{j \in S(i)} \neg\mathrm{Pr}(\overline{\neg\lambda_j}) \;\text{ and }\; \text{PA} \vdash \lambda_i \;\rightarrow\; \mathrm{Pr}(\overline{\bigvee_{j \in S(i)} \lambda_j}).$$

and furthermore $\lambda_i \rightarrow \neg\lambda_j$ for $i \neq j$ are constructed using *plural but finite* self-reference (solvability of n equations for n unknown sentences).

Self-referential sentences (continued)

Embedding a Kripke model to PA (Solovay)

Let $k = \langle W, R\rangle$, $W = \{1, \ldots, n\}$. Put $S(i) = \{\, j\,;\ i\,R\,j\,\}$. Example:



$S(1) = \{2, 3, 4\}$, $S(3) = \{4\}$, $S(2) = S(4) = \emptyset$.

Then consistent sentences $\lambda_1, \ldots, \lambda_n$ such that

$$\mathrm{PA} \vdash \lambda_i \rightarrow \textstyle\bigwedge_{j \in S(i)} \neg \mathrm{Pr}(\overline{\neg\lambda_j}) \ \text{ and } \ \mathrm{PA} \vdash \lambda_i \rightarrow \mathrm{Pr}(\overline{\textstyle\bigvee_{j \in S(i)} \lambda_j}).$$

and furthermore $\lambda_i \rightarrow \neg\lambda_j$ for $i \neq j$ are constructed using *plural but finite* self-reference (solvability of n equations for n unknown sentences).

A self-referential equation $\vdash \varphi \equiv \psi(\overline{\varphi})$ is *Gödelian* if ψ is built up using connectives and Pr only.

Provability logic GL

is a modal propositional logic with axioms and rules as follows:

A1: all propositional tautologies,

A2: $\Box(A \to B) \to (\Box A \to \Box B)$, MP: $A \to B, A \,/\, B$,

A3: $\Box A \to \Box\Box A$, Nec: $A \,/\, \Box A$.

A4: $\Box(\Box A \to A) \to \Box A$,

Arithmetic semantics of GL

Arithmetic valuation v is a function from modal formulas to sentences of arithmetic that preserves logical connectives and translates $\Box$ to Pr (atoms are sent to any sentences).

A formula A is a PA-*tautology* if $\mathrm{PA} \vdash v(A)$ for every translation v.

Provability logic GL

is a modal propositional logic with axioms and rules as follows:

A1: all propositional tautologies,

A2: $\Box(A \to B) \to (\Box A \to \Box B)$, MP: $A \to B, A \,/\, B$,

A3: $\Box A \to \Box\Box A$, Nec: $A \,/\, \Box A$.

A4: $\Box(\Box A \to A) \to \Box A$,

Arithmetic semantics of GL

Arithmetic valuation v is a function from modal formulas to sentences of arithmetic that preserves logical connectives and translates $\Box$ to Pr (atoms are sent to any sentences).
A formula A is a PA-*tautology* if PA $\vdash v(A)$ for every translation v.

Examples

$\Box p \to p$ is not a PA-tautology: for Gödel's sentence γ we have PA $\nvdash \mathrm{Pr}(\overline{\neg\gamma}) \to \neg\gamma$. A tautology: $\neg\Box\bot \to \neg\Box\neg\Box\bot$.

Properties and applications

1. Two completeness theorems: Kripke completeness w.r.t. finite transitive irreflexive frames, arithmetic completeness w.r.t. the semantics given above.

Properties and applications

1. Two completeness theorems: Kripke completeness w.r.t. finite transitive irreflexive frames, arithmetic completeness w.r.t. the semantics given above.
2. Every Gödelian equation has exactly one solution (up to provable equivalence).

Properties and applications

1. Two completeness theorems: Kripke completeness w.r.t. finite transitive irreflexive frames, arithmetic completeness w.r.t. the semantics given above.
2. Every Gödelian equation has exactly one solution (up to provable equivalence).
3. Explicit computability of the solutions:
   if PA $\vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$, then PA $\vdash \gamma \equiv \neg\mathrm{Con}(\mathsf{PA})$;
   if PA $\vdash \kappa \equiv \mathrm{Pr}(\overline{\kappa})$, then PA $\vdash \gamma \equiv 0 = 0$;
   if PA $\vdash \lambda \equiv \mathrm{Pr}(\overline{\lambda}) \rightarrow \kappa$, then PA $\vdash \lambda \equiv \mathrm{Pr}(\overline{\kappa}) \rightarrow \kappa$
   (can be verified by proving $\Box q \rightarrow q \equiv \Box(\Box q \rightarrow q) \rightarrow q$ in GL); etc.

Properties and applications

1. Two completeness theorems: Kripke completeness w.r.t. finite transitive irreflexive frames, arithmetic completeness w.r.t. the semantics given above.
2. Every Gödelian equation has exactly one solution (up to provable equivalence).
3. Explicit computability of the solutions:
   if PA $\vdash \gamma \equiv \mathrm{Pr}(\overline{\neg\gamma})$, then PA $\vdash \gamma \equiv \neg\mathrm{Con}(\mathsf{PA})$;
   if PA $\vdash \kappa \equiv \mathrm{Pr}(\overline{\kappa})$, then PA $\vdash \gamma \equiv 0 = 0$;
   if PA $\vdash \lambda \equiv \mathrm{Pr}(\overline{\lambda}) \rightarrow \kappa$, then PA $\vdash \lambda \equiv \mathrm{Pr}(\overline{\kappa}) \rightarrow \kappa$
   (can be verified by proving $\Box q \rightarrow q \equiv \Box(\Box q \rightarrow q) \rightarrow q$ in GL); etc.
4. Impossibility of symmetrically independent Gödelian sentence: no solution φ of a Gödelian equation $\vdash \varphi \equiv \psi(\overline{\varphi})$ satisfies PA $\vdash \mathrm{Con}(\mathsf{PA}) \rightarrow \neg\mathrm{Pr}(\overline{\varphi})\ \&\ \neg\mathrm{Pr}(\overline{\neg\varphi})$.

The logic R of Guaspari and Solovay

Besides $\Box$, we have "binary modalities" $\preceq, \prec$ in the modal language. These are applicable only to formulas starting with $\Box$. Thus $\Box\Box\bot \preceq \Box p$ is an example formula.

The logic R of Guaspari and Solovay

Besides $\Box$, we have "binary modalities" $\preceq, \prec$ in the modal language. These are applicable only to formulas starting with $\Box$. Thus $\Box\Box\bot \preceq \Box p$ is an example formula.

The theory R^- is obtained by adding to GL the axioms about witness comparison:

B1: $\Box A \preceq \Box B \;\rightarrow\; \Box A$,

B2: $\Box A \preceq \Box B \;\&\; \Box B \preceq \Box C \;\rightarrow\; \Box A \preceq \Box C$,

B3: $\Box A \prec \Box B \;\equiv\; \Box A \preceq \Box B \;\&\; \neg(\Box B \preceq \Box A)$,

B4: $\Box A \vee \Box B \;\rightarrow\; \Box A \preceq \Box B \vee \Box B \prec \Box A$,

P: $\Box A \preceq \Box B \;\rightarrow\; \Box(\Box A \preceq \Box B)$, $\quad \Box A \prec \Box B \;\rightarrow\; \Box(\Box A \prec \Box B)$.

The logic R of Guaspari and Solovay

Besides $\Box$, we have "binary modalities" $\preceq, \prec$ in the modal language. These are applicable only to formulas starting with $\Box$. Thus $\Box\Box\bot \preceq \Box p$ is an example formula.
The theory R^- is obtained by adding to GL the axioms about witness comparison:

B1: $\Box A \preceq \Box B \ \rightarrow\ \Box A$,

B2: $\Box A \preceq \Box B \ \&\ \Box B \preceq \Box C \ \rightarrow\ \Box A \preceq \Box C$,

B3: $\Box A \prec \Box B \ \equiv\ \Box A \preceq \Box B \ \&\ \neg(\Box B \preceq \Box A)$,

B4: $\Box A \vee \Box B \ \rightarrow\ \Box A \preceq \Box B \ \vee\ \Box B \prec \Box A$,

P: $\Box A \preceq \Box B \ \rightarrow\ \Box(\Box A \preceq \Box B)$, $\qquad \Box A \prec \Box B \ \rightarrow\ \Box(\Box A \prec \Box B)$.

The theory R has the additional rule $\Box A \,/\, A$.

The logic R of Guaspari and Solovay

Besides $\Box$, we have "binary modalities" $\preceq, \prec$ in the modal language. These are applicable only to formulas starting with $\Box$. Thus $\Box\Box\bot \preceq \Box p$ is an example formula.

The theory R^- is obtained by adding to GL the axioms about witness comparison:

B1: $\Box A \preceq \Box B \;\rightarrow\; \Box A$,

B2: $\Box A \preceq \Box B \;\&\; \Box B \preceq \Box C \;\rightarrow\; \Box A \preceq \Box C$,

B3: $\Box A \prec \Box B \;\equiv\; \Box A \preceq \Box B \;\&\; \neg(\Box B \preceq \Box A)$,

B4: $\Box A \vee \Box B \;\rightarrow\; \Box A \preceq \Box B \vee \Box B \prec \Box A$,

P: $\Box A \preceq \Box B \;\rightarrow\; \Box(\Box A \preceq \Box B)$, $\Box A \prec \Box B \;\rightarrow\; \Box(\Box A \prec \Box B)$.

The theory R has the additional rule $\Box A \,/\, A$.

Kripke semantics: formulas with $\preceq, \prec$ as the outermost symbol are treated as atoms (with the restrictions given by the axioms).

The logic R (continued)

Arithmetic semantics: the modalities $\preceq, \prec$ translate to $\preceq, \prec$:
$v(\Box A \preceq \Box B) = \mathrm{Pr}(\overline{v(A)}) \preceq \mathrm{Pr}(\overline{v(B)})$,
$v(\Box A \prec \Box B) = \mathrm{Pr}(\overline{v(A)}) \prec \mathrm{Pr}(\overline{v(B)})$.
However, the proof predicate and the corresponding provability predicate are let to vary.

The logic R (continued)

Arithmetic semantics: the modalities $\preceq, \prec$ translate to $\preceq, \prec$:
$v(\Box A \preceq \Box B) = \mathrm{Pr}(\overline{v(A)}) \preceq \mathrm{Pr}(\overline{v(B)})$,
$v(\Box A \prec \Box B) = \mathrm{Pr}(\overline{v(A)}) \prec \mathrm{Pr}(\overline{v(B)})$.
However, the proof predicate and the corresponding provability predicate are let to vary.

Completeness theorem

If R $\nvdash A$, then there exists a valuation v (i.e. a choice of a proof predicate and values of atoms) such that PA $\nvdash v(A)$.

Some applications of the logic R

1. R can prove Rosser's theorem in the form
$\Box(p \equiv \Box\neg p \preceq \Box p) \rightarrow (\neg\Box\bot \rightarrow (\neg\Box p \;\&\; \neg\Box\neg p))$.

Some applications of the logic R

1. R can prove Rosser's theorem in the form
   $\Box(p \equiv \Box\neg p \preceq \Box p) \to (\neg\Box\bot \to (\neg\Box p \;\&\; \neg\Box\neg p))$.
2. The Solovay's plural self-reference construction (concerning completeness of GL) can be formalized in GL (I believe).

Some applications of the logic R

1. R can prove Rosser's theorem in the form $\Box(p \equiv \Box\neg p \preceq \Box p) \to (\neg\Box\bot \to (\neg\Box p \ \& \ \neg\Box\neg p))$.
2. The Solovay's plural self-reference construction (concerning completeness of GL) can be formalized in GL (I believe).
3. Consider the formula $A = \Box(p \equiv \Box\neg p \preceq \Box p) \ \& \ \Box(q \equiv \Box\neg q \preceq \Box q) \to \Box(p \equiv q)$. Use Kripke semantics to show that this formula is not provable. Take the arithmetic counterexemple, i.e. sentences $v(p)$ and $v(q)$ and a proof predicate Pr such that $\mathbb{N} \models \mathrm{Pr}(\overline{p \equiv \mathrm{Pr}(\overline{\neg v(p)}) \preceq \mathrm{Pr}(\overline{v(p)})})$, $\mathbb{N} \models \mathrm{Pr}(\overline{q \equiv \mathrm{Pr}(\overline{\neg v(q)}) \preceq \mathrm{Pr}(\overline{v(q)})})$, but $\mathbb{N} \not\models \mathrm{Pr}(\overline{v(p) \equiv v(q)})$. Then $v(p)$ and $v(q)$ are non-equivalent.

Alternative Rosser logics. Why?

Look at the sentence of Hájková and Hájek: a number y such that $\neg\mathrm{Con}(\mathsf{PA}\restriction y + \overline{\mu})$ can be understood as a generalized proof of $\neg\mu$, and the formula $\neg\mathrm{Con}(\mathsf{PA}\restriction y + x)$ can be seen as a sort of proof predicate.

Alternative Rosser logics. Why?

Look at the sentence of Hájková and Hájek: a number y such that $\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y + \overline{\mu})$ can be understood as a generalized proof of $\neg\mu$, and the formula $\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y + x)$ can be seen as a sort of proof predicate. Then a sentence satisfying
$\mathsf{PA} \vdash \varphi \equiv \exists y(\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y + \overline{\varphi})\ \&\ \mathrm{Con}(\mathsf{PA}{\restriction}\, y + \overline{\varphi}))$
is in fact a Rosser's sentence.

Alternative Rosser logics. Why?

Look at the sentence of Hájková and Hájek: a number y such that $\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y+\overline{\mu})$ can be understood as a generalized proof of $\neg\mu$, and the formula $\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y+x)$ can be seen as a sort of proof predicate. Then a sentence satisfying

$\mathsf{PA} \vdash \varphi \equiv \exists y(\neg\mathrm{Con}(\mathsf{PA}{\restriction}\, y+\overline{\varphi})\ \&\ \mathrm{Con}(\mathsf{PA}{\restriction}\, y+\overline{\varphi}))$

is in fact a Rosser's sentence.

The logic Z has the axiom and the rule

W: $\Box A \rightarrow \Box(\neg B \rightarrow \Box A \prec \Box B)$, $\qquad A \,/\, \neg B \rightarrow \Box A \prec \Box B$

instead of the axiom P and the rule $\Box A \,/\, A$ of the theory R.

Alternative Rosser logics. Why?

Look at the sentence of Hájková and Hájek: a number y such that $\neg\mathrm{Con}(\mathsf{PA}\restriction y + \overline{\mu})$ can be understood as a generalized proof of $\neg\mu$, and the formula $\neg\mathrm{Con}(\mathsf{PA}\restriction y + x)$ can be seen as a sort of proof predicate. Then a sentence satisfying

$\mathsf{PA} \vdash \varphi \equiv \exists y(\neg\mathrm{Con}(\mathsf{PA}\restriction y + \overline{\varphi}) \ \&\ \mathrm{Con}(\mathsf{PA}\restriction y + \overline{\varphi}))$

is in fact a Rosser's sentence.

The logic Z has the axiom and the rule

W: $\Box A \rightarrow \Box(\neg B \rightarrow \Box A \prec \Box B)$, $\qquad A \,/\, \neg B \rightarrow \Box A \prec \Box B$

instead of the axiom P and the rule $\Box A \,/\, A$ of the theory R. This logic is arithmetically sound with respect to *all* proof predicates, i.e. standard in the sense of Guaspari and Solovay, or non-standard (like in the sentence of Hájková and Hájek).

Alternative Rosser logics. Why?

Look at the sentence of Hájková and Hájek: a number y such that $\neg\mathrm{Con}(\mathsf{PA}\restriction y+\overline{\mu})$ can be understood as a generalized proof of $\neg\mu$, and the formula $\neg\mathrm{Con}(\mathsf{PA}\restriction y+x)$ can be seen as a sort of proof predicate. Then a sentence satisfying

$\mathsf{PA}\vdash\varphi\equiv\exists y(\neg\mathrm{Con}(\mathsf{PA}\restriction y+\overline{\varphi})\ \&\ \mathrm{Con}(\mathsf{PA}\restriction y+\overline{\varphi}))$

is in fact a Rosser's sentence.

The logic Z has the axiom and the rule

W: $\Box A\rightarrow\Box(\neg B\rightarrow\Box A\prec\Box B)$, $\qquad A\,/\,\neg B\rightarrow\Box A\prec\Box B$

instead of the axiom P and the rule $\Box A\,/\,A$ of the theory R. This logic is arithmetically sound with respect to *all* proof predicates, i.e. standard in the sense of Guaspari and Solovay, or non-standard (like in the sentence of Hájková and Hájek). Example formula: $\Box\Box\bot\rightarrow\Box\Box\bot\prec\Box\bot$.