UNIFORMITY, CONTINGENCY, AND SELF-REFERENCE IN ARITHMETIC

Lingyuan Ye
joint work with Volker Halbach & Balthasar Grabmayr

2021.06.10

TsinghuaLogic Salon

Intro

Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet.
We thus have a sentence of arithmetic before us that statets its own unprovability.
— [Gödel, 1931, p.175]

When does an arithmetic sentence genuinely *express* some syntactical proposition?

When does an arithmetic *ascribe to itself* some syntactical proposition?

These questions concern the *intensionality* of arithmetic sentences.

> *Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet.*
> *We thus have a sentence of arithmetic before us that statets its own unprovability.*
> *— [Gödel, 1931, p.175]*

When does an arithmetic sentence genuinely *express* some syntactical proposition?

When does an arithmetic *ascribe to itself* some syntactical proposition?

These questions concern the *intensionality* of arithmetic sentences.

Intro

> *Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet.*
> *We thus have a sentence of arithmetic before us that statets its own unprovability.*
> *— [Gödel, 1931, p.175]*

When does an arithmetic sentence genuinely *express* some syntactical proposition?

When does an arithmetic *ascribe to itself* some syntactical proposition?

These questions concern the *intensionality* of arithmetic sentences.

Intro

> *Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet.*
> *We thus have a sentence of arithmetic before us that statets its own unprovability.*
> *— [Gödel, 1931, p.175]*

When does an arithmetic sentence genuinely *express* some syntactical proposition?

When does an arithmetic *ascribe to itself* some syntactical proposition?

These questions concern the *intensionality* of arithmetic sentences.

Intro

Self-reference is not trivial:

- For the proof of the First Incompleteness Theorem to work, we only need a sentence γ satisfying the following property,

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg \mathrm{Bew}(\ulcorner \gamma \urcorner)$$

- However, being a fixed-point in this sense is at best a *necessary* condition, far from a *sufficient*, for being self-referential.
- In particular, any provable sentence, e.g. $0 = 0$, is a fixed-point of $\mathrm{Bew}(x)$,

$$\mathrm{PA} \vdash 0 = 0 \leftrightarrow \mathrm{Bew}(\ulcorner 0 = 0 \urcorner)$$

- Does $0 = 0$, or any other provable sentence, *says of itself that it is provable*?

Intro

Self-reference is not trivial:

- For the proof of the First Incompleteness Theorem to work, we only need a sentence γ satisfying the following property,

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg \mathsf{Bew}(\ulcorner \gamma \urcorner)$$

- However, being a fixed-point in this sense is at best a *necessary* condition, far from a *sufficient*, for being self-referential.
- In particular, any provable sentence, e.g. $0 = 0$, is a fixed-point of $\mathsf{Bew}(x)$,

$$\mathrm{PA} \vdash 0 = 0 \leftrightarrow \mathsf{Bew}(\ulcorner 0 = 0 \urcorner)$$

- Does $0 = 0$, or any other provable sentence, *says of itself that it is provable*?

Intro

Self-reference is not trivial:

- For the proof of the First Incompleteness Theorem to work, we only need a sentence γ satisfying the following property,

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg \mathsf{Bew}(\ulcorner \gamma \urcorner)$$

- However, being a fixed-point in this sense is at best a *necessary* condition, far from a *sufficient*, for being self-referential.
- In particular, any provable sentence, e.g. $0 = 0$, is a fixed-point of $\mathsf{Bew}(x)$,

$$\mathrm{PA} \vdash 0 = 0 \leftrightarrow \mathsf{Bew}(\ulcorner 0 = 0 \urcorner)$$

- Does $0 = 0$, or any other provable sentence, *says of itself that it is provable*?

Intro

Self-reference is not trivial:

- For the proof of the First Incompleteness Theorem to work, we only need a sentence γ satisfying the following property,

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg\, \mathsf{Bew}(\ulcorner\gamma\urcorner)$$

- However, being a fixed-point in this sense is at best a *necessary* condition, far from a *sufficient*, for being self-referential.
- In particular, any provable sentence, e.g. $0 = 0$, is a fixed-point of $\mathsf{Bew}(x)$,

$$\mathrm{PA} \vdash 0 = 0 \leftrightarrow \mathsf{Bew}(\ulcorner 0 = 0 \urcorner)$$

- Does $0 = 0$, or any other provable sentence, *says of itself that it is provable*?

Intro

Self-reference is not trivial:

- For the proof of the First Incompleteness Theorem to work, we only need a sentence γ satisfying the following property,

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg \mathsf{Bew}(\ulcorner \gamma \urcorner)$$

- However, being a fixed-point in this sense is at best a *necessary* condition, far from a *sufficient*, for being self-referential.
- In particular, any provable sentence, e.g. $0 = 0$, is a fixed-point of $\mathsf{Bew}(x)$,

$$\mathrm{PA} \vdash 0 = 0 \leftrightarrow \mathsf{Bew}(\ulcorner 0 = 0 \urcorner)$$

- Does $0 = 0$, or any other provable sentence, *says of itself that it is provable*?

Intro

Why care about self-reference in arithmetic?

- Clear (massive) misconceptions about incompleteness results (self-reference, circularity, limits of intelligence ...);
- Pushforward technical development (cf. Löb's theorem, provability logic, Yablo's paradox, theory of truths ...);
- Evaluate philosophical implications of formal results.

Why care about self-reference in arithmetic?

- Clear (massive) misconceptions about incompleteness results (self-reference, circularity, limits of intelligence ...);
- Pushforward technical development (cf. Löb's theorem, provability logic, Yablo's paradox, theory of truths ...);
- Evaluate philosophical implications of formal results.

Intro

Why care about self-reference in arithmetic?

- Clear (massive) misconceptions about incompleteness results (self-reference, circularity, limits of intelligence ...);
- Pushforward technical development (cf. Löb's theorem, provability logic, Yablo's paradox, theory of truths ...);
- Evaluate philosophical implications of formal results.

Why care about self-reference in arithmetic?

- Clear (massive) misconceptions about incompleteness results (self-reference, circularity, limits of intelligence ...);
- Pushforward technical development (cf. Löb's theorem, provability logic, Yablo's paradox, theory of truths ...);
- Evaluate philosophical implications of formal results.

Henkin's Question [Henkin, 1952]

Is the sentence states its own provability provable or not?[1]

- [Kreisel, 1953] first proposed two constructions that were claimed to yield a provable and a refutable Henkin sentence.
- [Lob, 1955] gave his famous solution to Henkin's problem, according to which every sentence states its own provability is itself provable.
- Since then, provability logic, modalised fixed-point theorems, ...

[1] This is not Henkin's original formulation of his question, but this is what we believed Henkin had in mind.

Henkin's Question [Henkin, 1952]

Is the sentence states its own provability provable or not?[1]

- [Kreisel, 1953] first proposed two constructions that were claimed to yield a provable and a refutable Henkin sentence.
- [Lob, 1955] gave his famous solution to Henkin's problem, according to which every sentence states its own provability is itself provable.
- Since then, provability logic, modalised fixed-point theorems, ...

[1] This is not Henkin's original formulation of his question, but this is what we believed Henkin had in mind.

Henkin's Question [Henkin, 1952]

Is the sentence states its own provability provable or not?[1]

- [Kreisel, 1953] first proposed two constructions that were claimed to yield a provable and a refutable Henkin sentence.
- [Lob, 1955] gave his famous solution to Henkin's problem, according to which every sentence states its own provability is itself provable.
- Since then, provability logic, modalised fixed-point theorems, ...

[1] This is not Henkin's original formulation of his question, but this is what we believed Henkin had in mind.

Henkin's Question [Henkin, 1952]

Is the sentence states its own provability provable or not?[1]

- [Kreisel, 1953] first proposed two constructions that were claimed to yield a provable and a refutable Henkin sentence.
- [Lob, 1955] gave his famous solution to Henkin's problem, according to which every sentence states its own provability is itself provable.
- Since then, provability logic, modalised fixed-point theorems, ...

[1] This is not Henkin's original formulation of his question, but this is what we believed Henkin had in mind.

Theorem (Gödel's Second Incompleteness Theorem)

A sufficiently strong system can never prove its own consistency.

- The canonical consistency statement $\mathrm{Con}_{\mathrm{PA}}$ is not provable due to Gödel.
- But if we have used Rosser's provability predicate $\mathrm{Bew}^R(x)$,

$$\mathrm{Bew}^R(x) \equiv \exists y(B(y,x) \wedge \forall z < y \neg R(z,y))$$

then $\mathrm{Con}^R_{\mathrm{PA}} \equiv \neg \mathrm{Bew}^R(\ulcorner 0 = 1 \urcorner)$ would actually be provable.

Intro

Theorem (Gödel's Second Incompleteness Theorem)

A sufficiently strong system can never prove its own consistency.

- The canonical consistency statement $\mathrm{Con}_{\mathrm{PA}}$ is not provable due to Gödel.
- But if we have used Rosser's provability predicate $\mathrm{Bew}^R(x)$,

$$\mathrm{Bew}^R(x) \equiv \exists y(B(y,x) \land \forall z < y \neg R(z,y))$$

then $\mathrm{Con}^R_{\mathrm{PA}} \equiv \neg\, \mathrm{Bew}^R(\ulcorner 0 = 1 \urcorner)$ would actually be provable.

Theorem (Gödel's Second Incompleteness Theorem)

A sufficiently strong system can never prove its own consistency.

- The canonical consistency statement $\mathrm{Con}_{\mathrm{PA}}$ is not provable due to Gödel.
- But if we have used Rosser's provability predicate $\mathsf{Bew}^R(x)$,

$$\mathsf{Bew}^R(x) \equiv \exists y (B(y, x) \land \forall z < y \neg R(z, y))$$

then $\mathrm{Con}^R_{\mathrm{PA}} \equiv \neg \mathsf{Bew}^R(\ulcorner 0 = 1 \urcorner)$ would actually be provable.

Upshot: studying self-reference might lead to both philosophically and technically fruitful results.

Outline

Framework of Self-reference

Henkin’s Problem and Kreisel’s Construction

Contingency in Kreisel’s Construction

Solution by Uniformity

Framework of Self-reference

Sources of Intensionality

The following are basic steps to obtain self-referential sentences:

- Fix a *numbering* function α, mapping expressions (strings of symbols) and sequences of expressions to $\mathbb{N}$;
- Fix a *numeral* function ν, mapping $\mathbb{N}$ to closed terms in the language (the usual choice is simply $n \mapsto \overline{n}$). The composition $\nu \circ \alpha$ is the coding function $\ulcorner - \urcorner$;
- Choose a formula $B(x)$ that *expresses* the intended syntactic property P (according to $\ulcorner - \urcorner$);
- Construct a formula that ascribes to itself the property P via the chosen predicate $B(x)$.

Sources of Intensionality

The following are basic steps to obtain self-referential sentences:

- Fix a *numbering* function α, mapping expressions (strings of symbols) and sequences of expressions to $\mathbb{N}$;
- Fix a *numeral* function ν, mapping $\mathbb{N}$ to closed terms in the language (the usual choice is simply $n \mapsto \overline{n}$). The composition $\nu \circ \alpha$ is the coding function $\ulcorner - \urcorner$;
- Choose a formula $B(x)$ that *expresses* the intended syntactic property P (according to $\ulcorner - \urcorner$);
- Construct a formula that ascribes to itself the property P via the chosen predicate $B(x)$.

Sources of Intensionality

The following are basic steps to obtain self-referential sentences:

- Fix a *numbering* function α, mapping expressions (strings of symbols) and sequences of expressions to $\mathbb{N}$;
- Fix a *numeral* function ν, mapping $\mathbb{N}$ to closed terms in the language (the usual choice is simply $n \mapsto \overline{n}$). The composition $\nu \circ \alpha$ is the coding function $\ulcorner - \urcorner$;
- Choose a formula $B(x)$ that *expresses* the intended syntactic property P (according to $\ulcorner - \urcorner$);
- Construct a formula that ascribes to itself the property P via the chosen predicate $B(x)$.

Sources of Intensionality

The following are basic steps to obtain self-referential sentences:

- Fix a *numbering* function α, mapping expressions (strings of symbols) and sequences of expressions to $\mathbb{N}$;
- Fix a *numeral* function ν, mapping $\mathbb{N}$ to closed terms in the language (the usual choice is simply $n \mapsto \overline{n}$). The composition $\nu \circ \alpha$ is the coding function $\ulcorner - \urcorner$;
- Choose a formula $B(x)$ that *expresses* the intended syntactic property P (according to $\ulcorner - \urcorner$);
- Construct a formula that ascribes to itself the property P via the chosen predicate $B(x)$.

Sources of Intensionality

The following are basic steps to obtain self-referential sentences:

- Fix a *numbering* function α, mapping expressions (strings of symbols) and sequences of expressions to $\mathbb{N}$;
- Fix a *numeral* function ν, mapping $\mathbb{N}$ to closed terms in the language (the usual choice is simply $n \mapsto \overline{n}$). The composition $\nu \circ \alpha$ is the coding function $\ulcorner - \urcorner$;
- Choose a formula $B(x)$ that *expresses* the intended syntactic property P (according to $\ulcorner - \urcorner$);
- Construct a formula that ascribes to itself the property P via the chosen predicate $B(x)$.

Sources of Intensionality

In [Halbach and Visser, 2013], they characterises that each set corresponds to different *sources* of intensionality:

- [Grabmayr and Visser, 2020] has shown “crazy” numbering or numeral functions would lead to “crazy” results;
- Since [Feferman, 1960], various criteria have been proposed when does a predicate really *expresses* some property;
- [Halbach and Visser, 2013]also suggests that intensionality can be affected by how fixed-points are constructed.

Grabmayr, Halbach and myself are currently working on the last aspect, and how it interacts with other sources of intensionality. This will be the main topic of today.

Sources of Intensionality

In [Halbach and Visser, 2013], they characterises that each set corresponds to different *sources* of intensionality:

- [Grabmayr and Visser, 2020] has shown “crazy” numbering or numeral functions would lead to “crazy” results;
- Since [Feferman, 1960], various criteria have been proposed when does a predicate really *expresses* some property;
- [Halbach and Visser, 2013]also suggests that intensionality can be affected by how fixed-points are constructed.

Grabmayr, Halbach and myself are currently working on the last aspect, and how it interacts with other sources of intensionality. This will be the main topic of today.

Sources of Intensionality

In [Halbach and Visser, 2013], they characterises that each set corresponds to different *sources* of intensionality:

- [Grabmayr and Visser, 2020] has shown "crazy" numbering or numeral functions would lead to "crazy" results;
- Since [Feferman, 1960], various criteria have been proposed when does a predicate really *expresses* some property;
- [Halbach and Visser, 2013]also suggests that intensionality can be affected by how fixed-points are constructed.

Grabmayr, Halbach and myself are currently working on the last aspect, and how it interacts with other sources of intensionality. This will be the main topic of today.

Sources of Intensionality

In [Halbach and Visser, 2013], they characterises that each set corresponds to different *sources* of intensionality:

- [Grabmayr and Visser, 2020] has shown "crazy" numbering or numeral functions would lead to "crazy" results;
- Since [Feferman, 1960], various criteria have been proposed when does a predicate really *expresses* some property;
- [Halbach and Visser, 2013]also suggests that intensionality can be affected by how fixed-points are constructed.

Grabmayr, Halbach and myself are currently working on the last aspect, and how it interacts with other sources of intensionality. This will be the main topic of today.

Sources of Intensionality

In [Halbach and Visser, 2013], they characterises that each set corresponds to different *sources* of intensionality:

- [Grabmayr and Visser, 2020] has shown "crazy" numbering or numeral functions would lead to "crazy" results;
- Since [Feferman, 1960], various criteria have been proposed when does a predicate really *expresses* some property;
- [Halbach and Visser, 2013]also suggests that intensionality can be affected by how fixed-points are constructed.

Grabmayr, Halbach and myself are currently working on the last aspect, and how it interacts with other sources of intensionality. This will be the main topic of today.

Henkin's Problem and Kreisel's Construction

From now on we work in a sufficently strong and sound arithmetic language Σ (extends R); in particular, it contains some further function symbols other than $0, S, +, \times$; and we fix some coding $\ulcorner - \urcorner$.

Henkin and Kreisel was quite careful in expressing self-reference, since they do not want trivial fixed-points like $0 = 0$:

Kreisel-Henkin Criterion for Self-Reference

If $B(x)$ is given as expressing some property P, then a sentence *says of itself* it has property P, iff it is of the form $B(t)$ for some closed term t, such that

$$\Sigma \vdash t = \ulcorner B(t) \urcorner$$

We say t is a self-referential term w.r.t. $B(x)$.

Henkin and Kreisel was quite careful in expressing self-reference, since they do not want trivial fixed-points like $0 = 0$:

Kreisel-Henkin Criterion for Self-Reference

If $B(x)$ is given as expressing some property P, then a sentence *says of itself* it has property P, iff it is of the form $B(t)$ for some closed term t, such that

$$\Sigma \vdash t = \ulcorner B(t) \urcorner$$

We say t is a self-referential term w.r.t. $B(x)$.

Kreisel-Henkin Criterion

Let t be a self-referential term w.r.t. $B(x)$; suppose $B(x)$ expresses some property P.

- $B(t)$ intuitively expresses that the expression corresponds to t under the coding has property P;
- Since $\Sigma \vdash t = \ulcorner B(t) \urcorner$, the expression corresponds to t under the coding is simply $B(t)$;
- Hence, $B(t)$ expresses $B(t)$, viz. itself, has property P.

The sentence $B(t)$ is a fixed-point in the usual sense,

$$\Sigma \vdash B(t) \leftrightarrow B(\ulcorner B(t) \urcorner)$$

Kreisel-Henkin Criterion

Let t be a self-referential term w.r.t. $B(x)$; suppose $B(x)$ expresses some property P.

- $B(t)$ intuitively expresses that the expression corresponds to t under the coding has property P;
- Since $\Sigma \vdash t = \ulcorner B(t) \urcorner$, the expression corresponds to t under the coding is simply $B(t)$;
- Hence, $B(t)$ expresses $B(t)$, viz. itself, has property P.

The sentence $B(t)$ is a fixed-point in the usual sense,

$$\Sigma \vdash B(t) \leftrightarrow B(\ulcorner B(t) \urcorner)$$

Kreisel-Henkin Criterion

Let t be a self-referential term w.r.t. $B(x)$; suppose $B(x)$ expresses some property P.

- $B(t)$ intuitively expresses that the expression corresponds to t under the coding has property P;
- Since $\Sigma \vdash t = \ulcorner B(t) \urcorner$, the expression corresponds to t under the coding is simply $B(t)$;
- Hence, $B(t)$ expresses $B(t)$, viz. itself, has property P.

The sentence $B(t)$ is a fixed-point in the usual sense,

$$\Sigma \vdash B(t) \leftrightarrow B(\ulcorner B(t) \urcorner)$$

Kreisel-Henkin Criterion

Let t be a self-referential term w.r.t. $B(x)$; suppose $B(x)$ expresses some property P.

- $B(t)$ intuitively expresses that the expression corresponds to t under the coding has property P;
- Since $\Sigma \vdash t = \ulcorner B(t) \urcorner$, the expression corresponds to t under the coding is simply $B(t)$;
- Hence, $B(t)$ expresses $B(t)$, viz. itself, has property P.

The sentence $B(t)$ is a fixed-point in the usual sense,

$$\Sigma \vdash B(t) \leftrightarrow B(\ulcorner B(t) \urcorner)$$

Kreisel-Henkin Criterion

Let t be a self-referential term w.r.t. $B(x)$; suppose $B(x)$ expresses some property P.

- $B(t)$ intuitively expresses that the expression corresponds to t under the coding has property P;
- Since $\Sigma \vdash t = \ulcorner B(t) \urcorner$, the expression corresponds to t under the coding is simply $B(t)$;
- Hence, $B(t)$ expresses $B(t)$, viz. itself, has property P.

The sentence $B(t)$ is a fixed-point in the usual sense,

$$\Sigma \vdash B(t) \leftrightarrow B(\ulcorner B(t) \urcorner)$$

Gödel's Diagonal Construction

If Σ has enough function symbols, then the usual Gödel construction could yield such a term for any formula $\varphi(x)$.

Let $\mathrm{sub}(y, z)$ represent the function substituting a name z into the designated free variable x in the expression (coded by) y,

$$\Sigma \vdash \mathrm{sub}(\ulcorner\varphi(x)\urcorner, \ulcorner e\urcorner) = \ulcorner\varphi(\ulcorner e\urcorner)\urcorner$$

Then by definition we can prove

$$\Sigma \vdash \mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x, x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x, x))\urcorner) = \ulcorner\varphi(\mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x, x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x, x))\urcorner))\urcorner$$

Gödel's Diagonal Construction

If Σ has enough function symbols, then the usual Gödel construction could yield such a term for any formula $\varphi(x)$.

Let $\mathrm{sub}(y, z)$ represent the function substituting a name z into the designated free variable x in the expression (coded by) y,

$$\Sigma \vdash \mathrm{sub}(\ulcorner\varphi(x)\urcorner, \ulcorner e\urcorner) = \ulcorner\varphi(\ulcorner e\urcorner)\urcorner$$

Then by definition we can prove

$$\begin{aligned}\Sigma \vdash &\mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x,x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x,x))\urcorner) = \\ &\ulcorner\varphi(\mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x,x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x,x))\urcorner))\urcorner\end{aligned}$$

Gödel's Diagonal Construction

If Σ has enough function symbols, then the usual Gödel construction could yield such a term for any formula $\varphi(x)$.

Let $\mathrm{sub}(y, z)$ represent the function substituting a name z into the designated free variable x in the expression (coded by) y,

$$\Sigma \vdash \mathrm{sub}(\ulcorner\varphi(x)\urcorner, \ulcorner e\urcorner) = \ulcorner\varphi(\ulcorner e\urcorner)\urcorner$$

Then by definition we can prove

$$\begin{aligned}\Sigma \vdash &\mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x, x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x, x))\urcorner) = \\ &\quad \ulcorner\varphi(\mathrm{sub}(\ulcorner\varphi(\mathrm{sub}(x, x))\urcorner, \ulcorner\varphi(\mathrm{sub}(x, x))\urcorner))\urcorner\end{aligned}$$

Henkin's problem is then the following:

Henkin's Problem

Given a provability predicate $\mathrm{Bew}(x)$, if t is a self-referential term w.r.t. $\mathrm{Bew}(x)$, what's the provability-status of $\mathrm{Bew}(t)$?

We already know that if the standard provability predicate $\mathrm{Bew}(x)$ is used, then all these sentences will be provable, due to Löb's theorem.

However, Kreisel gives some other constructions that yields different results.

Henkin's problem is then the following:

Henkin's Problem

Given a provability predicate $\mathrm{Bew}(x)$, if t is a self-referential term w.r.t. $\mathrm{Bew}(x)$, what's the provability-status of $\mathrm{Bew}(t)$?

We already know that if the standard provability predicate $\mathrm{Bew}(x)$ is used, then all these sentences will be provable, due to Löb's theorem.

However, Kreisel gives some other constructions that yields different results.

Henkin's problem is then the following:

Henkin's Problem

Given a provability predicate $\mathrm{Bew}(x)$, if t is a self-referential term w.r.t. $\mathrm{Bew}(x)$, what's the provability-status of $\mathrm{Bew}(t)$?

We already know that if the standard provability predicate $\mathrm{Bew}(x)$ is used, then all these sentences will be provable, due to Löb's theorem.

However, Kreisel gives some other constructions that yields different results.

Of course, if Kreisel successfully obtained different results, he must have used other provability predicates:

Kreisel's Condition for Expressing Provability

A predicate $B(x)$ expresses provability iff it weakly represents provable sentences, i.e. for any sentence ψ,

$$\Sigma \vdash \psi \Leftrightarrow \Sigma \vdash B(\ulcorner \psi \urcorner)$$

The standard provability predicates $\mathrm{Bew}(x)$, Rosser's provability predicate $\mathrm{Bew}^R(x)$, and Kreisel's construction all satisfies this condition; only $\mathrm{Bew}(x)$ satisfies Löb's derivability condition.

Of course, if Kreisel successfully obtained different results, he must have used other provability predicates:

Kreisel's Condition for Expressing Provability

A predicate $B(x)$ expresses provability iff it weakly represents provable sentences, i.e. for any sentence ψ,

$$\Sigma \vdash \psi \Leftrightarrow \Sigma \vdash B(\ulcorner \psi \urcorner)$$

The standard provability predicates $\mathrm{Bew}(x)$, Rosser's provability predicate $\mathrm{Bew}^R(x)$, and Kreisel's construction all satisfies this condition; only $\mathrm{Bew}(x)$ satisfies Löb's derivability condition.

Kreisel's Construction

Kreisel's construction is as follows:

- Find a self-referential term t w.r.t. $x \neq x \wedge \mathrm{Bew}(x)$,

$$\Sigma \vdash t = \ulcorner t \neq t \wedge \mathrm{Bew}(t) \urcorner$$

- Define another predicate $\mathrm{Bew}^{\dagger}(x)$,

$$\mathrm{Bew}^{\dagger}(x) :\equiv x \neq t \wedge \mathrm{Bew}(x)$$

- t is a self-referential term w.r.t. $\mathrm{Bew}^{\dagger}(x)$,

$$\Sigma \vdash t = \ulcorner \mathrm{Bew}^{\dagger}(t) \urcorner$$

- $\mathrm{Bew}^{\dagger}(t) \equiv t \neq t \wedge \mathrm{Bew}(t)$ is refutable.

Kreisel's Construction

Kreisel's construction is as follows:

- Find a self-referential term t w.r.t. $x \neq x \wedge \mathsf{Bew}(x)$,

$$\Sigma \vdash t = \ulcorner t \neq t \wedge \mathsf{Bew}(t) \urcorner$$

- Define another predicate $\mathsf{Bew}^{\dagger}(x)$,

$$\mathsf{Bew}^{\dagger}(x) :\equiv x \neq t \wedge \mathsf{Bew}(x)$$

- t is a self-referential term w.r.t. $\mathsf{Bew}^{\dagger}(x)$,

$$\Sigma \vdash t = \ulcorner \mathsf{Bew}^{\dagger}(t) \urcorner$$

- $\mathsf{Bew}^{\dagger}(t) \equiv t \neq t \wedge \mathsf{Bew}(t)$ is refutable.

Kreisel's Construction

Kreisel's construction is as follows:

- Find a self-referential term t w.r.t. $x \neq x \wedge \mathsf{Bew}(x)$,

$$\Sigma \vdash t = \ulcorner t \neq t \wedge \mathsf{Bew}(t) \urcorner$$

- Define another predicate $\mathsf{Bew}^\dagger(x)$,

$$\mathsf{Bew}^\dagger(x) :\equiv x \neq t \wedge \mathsf{Bew}(x)$$

- t is a self-referential term w.r.t. $\mathsf{Bew}^\dagger(x)$,

$$\Sigma \vdash t = \ulcorner \mathsf{Bew}^\dagger(t) \urcorner$$

- $\mathsf{Bew}^\dagger(t) \equiv t \neq t \wedge \mathsf{Bew}(t)$ is refutable.

Kreisel's construction is as follows:

- Find a self-referential term t w.r.t. $x \neq x \wedge \mathsf{Bew}(x)$,

$$\Sigma \vdash t = \ulcorner t \neq t \wedge \mathsf{Bew}(t) \urcorner$$

- Define another predicate $\mathsf{Bew}^\dagger(x)$,

$$\mathsf{Bew}^\dagger(x) :\equiv x \neq t \wedge \mathsf{Bew}(x)$$

- t is a self-referential term w.r.t. $\mathsf{Bew}^\dagger(x)$,

$$\Sigma \vdash t = \ulcorner \mathsf{Bew}^\dagger(t) \urcorner$$

- $\mathsf{Bew}^\dagger(t) \equiv t \neq t \wedge \mathsf{Bew}(t)$ is refutable.

Kreisel's Construction

Kreisel's construction is as follows:

- Find a self-referential term t w.r.t. $x \neq x \wedge \mathsf{Bew}(x)$,

$$\Sigma \vdash t = \ulcorner t \neq t \wedge \mathsf{Bew}(t) \urcorner$$

- Define another predicate $\mathsf{Bew}^{\dagger}(x)$,

$$\mathsf{Bew}^{\dagger}(x) :\equiv x \neq t \wedge \mathsf{Bew}(x)$$

- t is a self-referential term w.r.t. $\mathsf{Bew}^{\dagger}(x)$,

$$\Sigma \vdash t = \ulcorner \mathsf{Bew}^{\dagger}(t) \urcorner$$

- $\mathsf{Bew}^{\dagger}(t) \equiv t \neq t \wedge \mathsf{Bew}(t)$ is refutable.

Kreisel's Construction

It remains to show that $\mathsf{Bew}^\dagger(x)$ is a provability predicate in Kreisel's sense:

$\mathsf{Bew}^\dagger(x)$ weakly represents provability.

Given any sentence ψ, we distinguish two cases:

1. If ψ is not $\mathsf{Bew}^\dagger(t)$, viz. $t \neq t \wedge \mathsf{Bew}(t)$, then $\Sigma \vdash t \neq \ulcorner\psi\urcorner$, so

$$\Sigma \vdash \mathsf{Bew}^\dagger(\ulcorner\psi\urcorner) \Leftrightarrow \Sigma \vdash \mathsf{Bew}(\ulcorner\psi\urcorner) \Leftrightarrow \Sigma \vdash \psi$$

2. If ψ is $\mathsf{Bew}^\dagger(t)$, then $\Sigma \vdash t = \ulcorner\psi\urcorner$,

$$\Sigma \vdash \mathsf{Bew}^\dagger(\ulcorner\psi\urcorner) \leftrightarrow \mathsf{Bew}^\dagger(t) \leftrightarrow \psi$$

Hence $\mathsf{Bew}^\dagger(x)$ indeed weakly represents provability.

$\square$

Kreisel's Construction

It remains to show that $\mathsf{Bew}^{\dagger}(x)$ is a provability predicate in Kreisel's sense:

$\mathsf{Bew}^{\dagger}(x)$ weakly represents provability.

Given any sentence ψ, we distinguish two cases:

1. If ψ is not $\mathsf{Bew}^{\dagger}(t)$, viz. $t \neq t \wedge \mathsf{Bew}(t)$, then $\Sigma \vdash t \neq \ulcorner\psi\urcorner$, so

$$\Sigma \vdash \mathsf{Bew}^{\dagger}(\ulcorner\psi\urcorner) \Leftrightarrow \Sigma \vdash \mathsf{Bew}(\ulcorner\psi\urcorner) \Leftrightarrow \Sigma \vdash \psi$$

2. If ψ is $\mathsf{Bew}^{\dagger}(t)$, then $\Sigma \vdash t = \ulcorner\psi\urcorner$,

$$\Sigma \vdash \mathsf{Bew}^{\dagger}(\ulcorner\psi\urcorner) \leftrightarrow \mathsf{Bew}^{\dagger}(t) \leftrightarrow \psi$$

Hence $\mathsf{Bew}^{\dagger}(x)$ indeed weakly represents provability.

$\square$

Kreisel's Construction

We now have a *refutable* sentence that claims its own provability.

- Literatures usually argue that $\mathrm{Bew}^\dagger(x)$ is not a good provability predicate, hence it does not genuinely express provability.
- [Smoryński, 2008, Halbach and Visser, 2013] points out the way Kreisel constructs the self-referential term t of $\mathrm{Bew}^\dagger(x)$ is also not standard.
- More interestingly, the fact is, if we apply Gödel's construction to the predicate $\mathrm{Bew}^\dagger(x)$ and obtain another self-referential term t' w.r.t. $\mathrm{Bew}^\dagger(x)$, we would obtain a *provable* Henkin sentence $\mathrm{Bew}^\dagger(t')$.
- Are $\mathrm{Bew}^\dagger(t)$ and $\mathrm{Bew}^\dagger(t')$ philosophically different, at least for their intensional aspects?

Kreisel's Construction

We now have a *refutable* sentence that claims its own provability.

- Literatures usually argue that $\mathrm{Bew}^{\dagger}(x)$ is not a good provability predicate, hence it does not genuinely express provability.
- [Smoryński, 2008, Halbach and Visser, 2013] points out the way Kreisel constructs the self-referential term t of $\mathrm{Bew}^{\dagger}(x)$ is also not standard.
- More interestingly, the fact is, if we apply Gödel's construction to the predicate $\mathrm{Bew}^{\dagger}(x)$ and obtain another self-referential term t' w.r.t. $\mathrm{Bew}^{\dagger}(x)$, we would obtain a *provable* Henkin sentence $\mathrm{Bew}^{\dagger}(t')$.
- Are $\mathrm{Bew}^{\dagger}(t)$ and $\mathrm{Bew}^{\dagger}(t')$ philosophically different, at least for their intensional aspects?

Kreisel's Construction

We now have a *refutable* sentence that claims its own provability.

- Literatures usually argue that $\mathrm{Bew}^{\dagger}(x)$ is not a good provability predicate, hence it does not genuinely express provability.
- [Smoryński, 2008, Halbach and Visser, 2013] points out the way Kreisel constructs the self-referential term t of $\mathrm{Bew}^{\dagger}(x)$ is also not standard.
- More interestingly, the fact is, if we apply Gödel's construction to the predicate $\mathrm{Bew}^{\dagger}(x)$ and obtain another self-referential term t' w.r.t. $\mathrm{Bew}^{\dagger}(x)$, we would obtain a *provable* Henkin sentence $\mathrm{Bew}^{\dagger}(t')$.
- Are $\mathrm{Bew}^{\dagger}(t)$ and $\mathrm{Bew}^{\dagger}(t')$ philosophically different, at least for their intensional aspects?

Kreisel's Construction

We now have a *refutable* sentence that claims its own provability.

- Literatures usually argue that $\mathrm{Bew}^{\dagger}(x)$ is not a good provability predicate, hence it does not genuinely express provability.
- [Smoryński, 2008, Halbach and Visser, 2013] points out the way Kreisel constructs the self-referential term t of $\mathrm{Bew}^{\dagger}(x)$ is also not standard.
- More interestingly, the fact is, if we apply Gödel's construction to the predicate $\mathrm{Bew}^{\dagger}(x)$ and obtain another self-referential term t' w.r.t. $\mathrm{Bew}^{\dagger}(x)$, we would obtain a *provable* Henkin sentence $\mathrm{Bew}^{\dagger}(t')$.
- Are $\mathrm{Bew}^{\dagger}(t)$ and $\mathrm{Bew}^{\dagger}(t')$ philosophically different, at least for their intensional aspects?

Kreisel's Construction

We now have a *refutable* sentence that claims its own provability.

- Literatures usually argue that $\mathsf{Bew}^\dagger(x)$ is not a good provability predicate, hence it does not genuinely express provability.
- [Smoryński, 2008, Halbach and Visser, 2013] points out the way Kreisel constructs the self-referential term t of $\mathsf{Bew}^\dagger(x)$ is also not standard.
- More interestingly, the fact is, if we apply Gödel's construction to the predicate $\mathsf{Bew}^\dagger(x)$ and obtain another self-referential term t' w.r.t. $\mathsf{Bew}^\dagger(x)$, we would obtain a *provable* Henkin sentence $\mathsf{Bew}^\dagger(t')$.
- Are $\mathsf{Bew}^\dagger(t)$ and $\mathsf{Bew}^\dagger(t')$ philosophically different, at least for their intensional aspects?

Contingency in Kreisel's Construction

Contingency I — Prediagonalisation

Let's examine *how* the self-referential term is obtained in Kreisel's construction:

- Usually, to construct a self-referential term of some formula $B(x)$, we apply some diagonal operator to it (e.g. Gödel's construction described before);
- However, Kreisel's cosntruction is not like that; We first diagonalise $x \neq x \wedge \mathrm{Bew}(x)$ and obtain t, and t automatically becomes a self-referential term of $\mathrm{Bew}^{\dagger}(x)$;
- This is a form of *prediagonalisation*. Moreover, t is also a self-referential term w.r.t. the following very different sentence,

$$x \neq x \wedge \mathrm{Bew}^{\dagger}(t)$$

Contingency I — Prediagonalisation

Let's examine *how* the self-referential term is obtained in Kreisel's construction:

- Usually, to construct a self-referential term of some formula $B(x)$, we apply some diagonal operator to it (e.g. Gödel's construction described before);
- However, Kreisel's cosntruction is not like that; We first diagonalise $x \neq x \wedge \mathrm{Bew}(x)$ and obtain t, and t automatically becomes a self-referential term of $\mathrm{Bew}^{\dagger}(x)$;
- This is a form of *prediagonalisation*. Moreover, t is also a self-referential term w.r.t. the following very different sentence,

$$x \neq x \wedge \mathrm{Bew}^{\dagger}(t)$$

Contingency I — Prediagonalisation

Let's examine *how* the self-referential term is obtained in Kreisel's construction:

- Usually, to construct a self-referential term of some formula $B(x)$, we apply some diagonal operator to it (e.g. Gödel's construction described before);
- However, Kreisel's cosntruction is not like that; We first diagonalise $x \neq x \wedge \mathrm{Bew}(x)$ and obtain t, and t automatically becomes a self-referential term of $\mathrm{Bew}^{\dagger}(x)$;
- This is a form of *prediagonalisation*. Moreover, t is also a self-referential term w.r.t. the following very different sentence,

$$x \neq x \wedge \mathrm{Bew}^{\dagger}(t)$$

Contingency I — Prediagonalisation

Let's examine *how* the self-referential term is obtained in Kreisel's construction:

- Usually, to construct a self-referential term of some formula $B(x)$, we apply some diagonal operator to it (e.g. Gödel's construction described before);
- However, Kreisel's cosntruction is not like that; We first diagonalise $x \neq x \wedge \mathsf{Bew}(x)$ and obtain t, and t automatically becomes a self-referential term of $\mathsf{Bew}^{\dagger}(x)$;
- This is a form of *prediagonalisation.* Moreover, t is also a self-referential term w.r.t. the following very different sentence,

$$x \neq x \wedge \mathsf{Bew}^{\dagger}(t)$$

More generally is the following observation:

Observation

Let $\varphi(x, \cdots, x)$ be a formula with n (marked) free occurrences of x. Suppose we have obtained a closed term t

$$\Sigma \vdash t = \ulcorner \varphi(t) \urcorner$$

For any strict subset S of $\{1, \cdots, n\}$, we can obtain a formula $\varphi_S(x)$ by substituting t into the i-th occurrences of x in φ, whenever $i \in S$. Automatically, for any such S,

$$\Sigma \vdash t = \ulcorner \varphi_S(t) \urcorner$$

The number of such different formulas is $2^n - 1$!

More generally is the following observation:

Observation

Let $\varphi(x, \cdots, x)$ be a formula with n (marked) free occurrences of x. Suppose we have obtained a closed term t

$$\Sigma \vdash t = \ulcorner \varphi(t) \urcorner$$

For any strict subset S of $\{1, \cdots, n\}$, we can obtain a formula $\varphi_S(x)$ by substituting t into the i-th occurrences of x in φ, whenever $i \in S$. Automatically, for any such S,

$$\Sigma \vdash t = \ulcorner \varphi_S(t) \urcorner$$

The number of such different formulas is $2^n - 1$!

Contingency I — Prediagonalisation

More generally is the following observation:

Observation

Let $\varphi(x, \cdots, x)$ be a formula with n (marked) free occurrences of x. Suppose we have obtained a closed term t

$$\Sigma \vdash t = \ulcorner \varphi(t) \urcorner$$

For any strict subset S of $\{1, \cdots, n\}$, we can obtain a formula $\varphi_S(x)$ by substituting t into the i-th occurrences of x in φ, whenever $i \in S$. Automatically, for any such S,

$$\Sigma \vdash t = \ulcorner \varphi_S(t) \urcorner$$

The number of such different formulas is $2^n - 1$!

In Natural Language?

Why no such absurd consequences in natural language?

- We generally avoid the use of multiple occurrences of variables:

$$x \text{ is } \textit{both} \text{ provable } \textit{and} \text{ unprovable.}$$

- Formalise in arithmetic, we have multiple occurrences of the variable,

$$\mathsf{Bew}(x) \wedge \neg \mathsf{Bew}(x)$$

In Natural Language?

Why no such absurd consequences in natural language?

- We generally avoid the use of multiple occurrences of variables:

  x is *both* provable *and* unprovable.

- Formalise in arithmetic, we have multiple occurrences of the variable,

$$\mathrm{Bew}(x) \wedge \neg \mathrm{Bew}(x)$$

In Natural Language?

Why no such absurd consequences in natural language?

- We generally avoid the use of multiple occurrences of variables:

  x is *both* provable *and* unprovable.

- Formalise in arithmetic, we have multiple occurrences of the variable,

$$\mathsf{Bew}(x) \wedge \neg \mathsf{Bew}(x)$$

In Natural Language?

Why no such absurd consequences in natural language?

- When formalise self-reference, in natural language

  This sentence is *both* provable *and* unprovable.

- Formalise in arithmetic,

$$\mathsf{Bew}(t) \wedge \neg \mathsf{Bew}(t)$$

  This makes t self-referential also w.r.t.

$$\mathsf{Bew}(t) \wedge \neg \mathsf{Bew}(x), \; \mathsf{Bew}(x) \wedge \neg \mathsf{Bew}(t)$$

In Natural Language?

Why no such absurd consequences in natural language?

- When formalise self-reference, in natural language

  This sentence is *both* provable *and* unprovable.

- Formalise in arithmetic,

$$\mathsf{Bew}(t) \wedge \neg\, \mathsf{Bew}(t)$$

  This makes t self-referential also w.r.t.

$$\mathsf{Bew}(t) \wedge \neg\, \mathsf{Bew}(x),\ \mathsf{Bew}(x) \wedge \neg\, \mathsf{Bew}(t)$$

Contingency II — Functional v.s. Accidental

[Van Fraassen, 1970] has distinguished *functional* and *accidental* self-reference.

Consider the following two sentences:

$$\text{This sentence is false.} \tag{A}$$

$$\text{The sentence written on my blackboard is false.} \tag{A1}$$

Contingency II — Functional v.s. Accidental

Accidental self-reference are very "dangerous" in arithmetic, because, unlike natural language, you lack *canonical choices.*

- In natural language, indexicals can induce paradigmatic self-reference due to its functionality.
- Kreisel's construction is in particular *not* functional; it depends on very specific features of the formula being diagonalised.
- *Uniformity* is supposed to provide functionality in self-reference.

Contingency II — Functional v.s. Accidental

Accidental self-reference are very "dangerous" in arithmetic, because, unlike natural language, you lack *canonical choices.*

- In natural language, indexicals can induce paradigmatic self-reference due to its functionality.
- Kreisel's construction is in particular *not* functional; it depends on very specific features of the formula being diagonalised.
- *Uniformity* is supposed to provide functionality in self-reference.

Contingency II — Functional v.s. Accidental

Accidental self-reference are very "dangerous" in arithmetic, because, unlike natural language, you lack *canonical choices.*

- In natural language, indexicals can induce paradigmatic self-reference due to its functionality.
- Kreisel's construction is in particular *not* functional; it depends on very specific features of the formula being diagonalised.
- *Uniformity* is supposed to provide functionality in self-reference.

Contingency II — Functional v.s. Accidental

Accidental self-reference are very "dangerous" in arithmetic, because, unlike natural language, you lack *canonical choices*.

- In natural language, indexicals can induce paradigmatic self-reference due to its functionality.
- Kreisel's construction is in particular *not* functional; it depends on very specific features of the formula being diagonalised.
- *Uniformity* is supposed to provide functionality in self-reference.

Solution by Uniformity

Uniformity was first brought up in [Halbach and Visser, 2013], however, the definition there does not work.

What do we want for uniform diagonalisation?

- It should yield fixed-points for every formula with a designated variable;
- The allowed construction should be able to be carried out in a reasonable *syntax theory* (cf. [Halbach and Leigh, 2021]).

Uniform Diagonalisation

Uniformity was first brought up in [Halbach and Visser, 2013], however, the definition there does not work.

What do we want for uniform diagonalisation?

- It should yield fixed-points for every formula with a designated variable;
- The allowed construction should be able to be carried out in a reasonable *syntax theory* (cf. [Halbach and Leigh, 2021]).

Uniform Diagonalisation

Uniformity was first brought up in [Halbach and Visser, 2013], however, the definition there does not work.

What do we want for uniform diagonalisation?

- It should yield fixed-points for every formula with a designated variable;
- The allowed construction should be able to be carried out in a reasonable *syntax theory* (cf. [Halbach and Leigh, 2021]).

Uniform Diagonalisation

Uniformity was first brought up in [Halbach and Visser, 2013], however, the definition there does not work.

What do we want for uniform diagonalisation?

- It should yield fixed-points for every formula with a designated variable;
- The allowed construction should be able to be carried out in a reasonable *syntax theory* (cf. [Halbach and Leigh, 2021]).

Uniform Diagonalisation

We consider formulas and terms with free variable x. Allowed constructions:

1. Two meta-linguistic substitution functions:

$$\mathsf{Sub}_f : \mathsf{Fml}_x \times \mathsf{Term}_x \to \mathsf{Fml}_x$$

$$\mathsf{Sub}_t : \mathsf{Term}_x \times \mathsf{Term}_x \to \mathsf{Term}_x$$

2. The naming functions for formulas and terms:

$$\ulcorner - \urcorner_f : \mathsf{Fml}_x \to \mathsf{Term}_x$$

$$\ulcorner - \urcorner_t : \mathsf{Term}_x \to \mathsf{Term}_x$$

3. Given any n-ary logical connective $\star$, the following meta-linguistic function:

$$\overline{\star} : \mathsf{Fml}_x^n \to \mathsf{Fml}_x$$

4. Given any n-ary function symbol f, the following meta-linguistic function:

$$\overline{f} : \mathsf{Term}_x^n \to \mathsf{Term}_x$$

5. Given any n-ary predicate symbol R, the following meta-linguistic function:

$$\overline{R} : \mathsf{Term}_x^n \to \mathsf{Fml}_x$$

6. Also, we should be able to get access to the variable term x,

$$\overline{x} : 1 \to \mathsf{Term}_x$$

   where 1 is the singleton set, and $\overline{x}$ picks out the term x in Term_x.

Uniform Diagonalisation

Definition

A uniform diagonal operator is a function

$$d : \mathsf{Fml}_x \to \mathsf{Term}_x$$

which is constructed by finitely applying the above operations, such that for any formula $\varphi(x)$,

$$\Sigma \vdash d\varphi = \ulcorner \varphi(d\varphi) \urcorner$$

Gödel's Construction Is Uniform

Gödel's construction can be viewed as the following composition,

$$\mathsf{Fml}_x \xrightarrow{\cong} \mathsf{Fml}_x \times 1 \xrightarrow{\mathsf{id} \times \overline{\mathrm{sub}(x,x)}} \mathsf{Fml}_x \times \mathsf{Term}_x \xrightarrow{\mathsf{Sub}_f} \mathsf{Fml}_x \xrightarrow{\ulcorner - \urcorner_f}$$

$$\mathsf{Term}_x \xrightarrow{\langle \mathsf{id}, \mathsf{id} \rangle} \mathsf{Term}_x \times \mathsf{Term}_x \xrightarrow{\overline{\mathrm{sub}}} \mathsf{Term}_x$$

Given any formula $\varphi(x)$,

$$\begin{aligned}\varphi(x) \mapsto& (\varphi(x), *) \mapsto (\varphi(x), \mathrm{sub}(x,x)) \mapsto \varphi(\mathrm{sub}(x,x)) \\ &\mapsto \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner \mapsto (\ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner, \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner) \\ &\mapsto \mathrm{sub}(\ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner, \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner)\end{aligned}$$

Gödel's Construction Is Uniform

Gödel's construction can be viewed as the following composition,

$$\mathsf{Fml}_x \xrightarrow{\cong} \mathsf{Fml}_x \times 1 \xrightarrow{\mathsf{id} \times \overline{\mathrm{sub}(x,x)}} \mathsf{Fml}_x \times \mathsf{Term}_x \xrightarrow{\mathsf{Sub}_f} \mathsf{Fml}_x \xrightarrow{\ulcorner - \urcorner_f} \mathsf{Term}_x \xrightarrow{\langle \mathsf{id}, \mathsf{id} \rangle} \mathsf{Term}_x \times \mathsf{Term}_x \xrightarrow{\overline{\mathrm{sub}}} \mathsf{Term}_x$$

Given any formula $\varphi(x)$,

$$\begin{aligned}\varphi(x) &\mapsto (\varphi(x), *) \mapsto (\varphi(x), \mathrm{sub}(x,x)) \mapsto \varphi(\mathrm{sub}(x,x)) \\ &\mapsto \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner \mapsto (\ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner, \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner) \\ &\mapsto \mathrm{sub}(\ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner, \ulcorner \varphi(\mathrm{sub}(x,x)) \urcorner)\end{aligned}$$

Kreisel-Like Construction

Our goal is to rule out accidental fixed-point constructions like Kreisel-like ones.

What we want is for the following conjecture to be true:

Conjecture

Let $\varphi(x, x)$ be a formula with two marked strings of a designated variable x. For any closed term t, it cannot be obtained by applying a uniform diagonal operator to the formula $\varphi(t, x)$.

Our goal is to rule out accidental fixed-point constructions like Kreisel-like ones.

What we want is for the following conjecture to be true:

Conjecture

Let $\varphi(x, x)$ be a formula with two marked strings of a designated variable x. For any closed term t, it cannot be obtained by applying a uniform diagonal operator to the formula $\varphi(t, x)$.

Kreisel-Like Construction

However, the Conjecture does not hold for all naming functions:

- Using the standard numeral function $n \mapsto \overline{n}$, we can construct a contrived numbering function α s.t. Kreisel-like fixed-points are possible to be obtained uniformly;
- Moreover, using the standard numbering function (cf. [Gödel, 1931]), we can construct a contrivied numeral function ν s.t. Kreisel-like fixed-points are possible to be obtained uniformly.

Upshot: it is important to constrain *both* the numbering *and* the numeral function.

Kreisel-Like Construction

However, the Conjecture does not hold for all naming functions:

- Using the standard numeral function $n \mapsto \overline{n}$, we can construct a contrived numbering function α s.t. Kreisel-like fixed-points are possible to be obtained uniformly;
- Moreover, using the standard numbering function (cf. [Gödel, 1931]), we can construct a contrivied numeral function ν s.t. Kreisel-like fixed-points are possible to be obtained uniformly.

Upshot: it is important to constrain *both* the numbering *and* the numeral function.

Kreisel-Like Construction

However, the Conjecture does not hold for all naming functions:

- Using the standard numeral function $n \mapsto \overline{n}$, we can construct a contrived numbering function α s.t. Kreisel-like fixed-points are possible to be obtained uniformly;
- Moreover, using the standard numbering function (cf. [Gödel, 1931]), we can construct a contrivied numeral function ν s.t. Kreisel-like fixed-points are possible to be obtained uniformly.

Upshot: it is important to constrain *both* the numbering *and* the numeral function.

Kreisel-Like Construction

However, the Conjecture does not hold for all naming functions:

- Using the standard numeral function $n \mapsto \overline{n}$, we can construct a contrived numbering function α s.t. Kreisel-like fixed-points are possible to be obtained uniformly;
- Moreover, using the standard numbering function (cf. [Gödel, 1931]), we can construct a contrivied numeral function ν s.t. Kreisel-like fixed-points are possible to be obtained uniformly.

Upshot: it is important to constrain *both* the numbering *and* the numeral function.

Well-founded Naming

Definition

Let $\preceq$ denote the subexpression relation. We say an expression e is **weakly named** in another expression e', denoted as $e \blacktriangleleft e'$, if there exists another expression e'' such that

$$e \preceq e'' \ \& \ulcorner e'' \urcorner \preceq e'$$

Lemma

If t is a self-referential term obtained by applying a uniform diagonal operator d to the formula $\varphi(x)$, then there must exists a term s that

$$\varphi(s) \blacktriangleleft t$$

Definition

Let $\preceq$ denote the subexpression relation. We say an expression e is **weakly named** in another expression e', denoted as $e \blacktriangleleft e'$, if there exists another expression e'' such that

$$e \preceq e'' \,\&\, \ulcorner e'' \urcorner \preceq e'$$

Lemma

If t is a self-referential term obtained by applying a uniform diagonal operator d to the formula $\varphi(x)$, then there must exists a term s that

$$\varphi(s) \blacktriangleleft t$$

Well-founded Naming

Theorem

When the weakly naming relation $\blacktriangleleft$ is well-founded, the Conjecture does hold.

Proof.

If t is obtained uniformly from $\varphi(t, x)$, then by the previous lemma, we know for some term s,

$$\varphi(t, s) \blacktriangleleft t$$

However, this implies that

$$t \blacktriangleleft t$$

contradicting the well-founded assumption. □

Theorem

When the weakly naming relation $\blacktriangleleft$ is well-founded, the Conjecture does hold.

Proof.

If t is obtained uniformly from $\varphi(t, x)$, then by the previous lemma, we know for some term s,

$$\varphi(t, s) \blacktriangleleft t$$

However, this implies that

$$t \blacktriangleleft t$$

contradicting the well-founded assumption. $\square$

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

What We Have Achieved

Some concluding remarks:

- We've proved Kreisel-like construction is definitely not uniform;
- Technically there's much more going on (cf. our draft);
- If our intuition about the Gödel naming is like *quotation*, then we expect the weakly naming function to be *well-founded*; meta-mathematically justify quotation reading of coding;
- Uniformity further shows the sensitivity of intensional issues in a technical context;
- The current framework can be easily generalised to other constructions, leading us further towards true intensionality in arithmetic.

Feferman, S. (1960).
Arithmetization of metamathematics in a general setting.
Fundamenta mathematicae, 49(1):35–92.

Gödel, K. (1931).
Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i.
Monatshefte für mathematik und physik, 38(1):173–198.

Grabmayr, B. and Visser, A. (2020).
Self-reference upfront: A study of self-referential gödel numberings.
arXiv preprint arXiv:2006.12178.

Halbach, V. and Leigh, G. (2021).
The road to paradox: A guide to syntax, truth, and modality.
To be published.

Halbach, V. and Visser, A. (2013).
Self-reference in arithmetic.
Logic Group Preprint Series, 316:1–59.

Henkin, L. (1952).
A problem concerning provability.
Journal of Symbolic Logic, (17):160.

Kreisel, G. (1953).
On a problem of henkin's.
Indagationes Mathematicae, (15):405–406.

Lob, M. H. (1955).
Solution of a problem of leon henkin.
The Journal of Symbolic Logic, 20(2):115–118.

Smoryński, C. (2008).
***The development of self-reference: Löb's theorem*, pages 110–133.**
Springer.

Van Fraassen, B. C. (1970).
Inference and self-reference.
Synthese, pages 425–438.

Thanks for Listening!

Q & A