Self-reference and intensionality in metamathematics

Volker Halbach

16 August 2021

International Workshop on Gödel's Incompleteness Theorems
Wuhan

The purpose of the talk

This is mainly a programmatic overview of some work I have done with others.

The technical results have or will be published here:

‘Self-Reference in Arithmetic I’ (with Albert Visser), *Review of Symbolic Logic* 7 (2014), 671–691

‘Self-Reference in Arithmetic II’ (with Albert Visser), *Review of Symbolic Logic* 7 (2014), 692–712

‘The Henkin sentence’ (with Albert Visser), in *The Life and Work of Leon Henkin* (*Essays on His Contributions*), María Manzano, Ildiko Sain and Enrique Alonso (eds.), Studies in Universal Logic, Birkhäuser, Basel, 2014, 249–264

The Road to Paradox: A Guide to Syntax, Truth, and 15 Modality, (with Graham Leigh), Cambridge University Press, almost finished

Contingency, uniformity and well-founded naming with Balthasar Grabmayr and Lingyuan Ye

On the Transubstantiation of Words Into Numbers with Balthasar Grabmayr, Beau Mount, and Albert Visser

The plan

Philosophical significance

Intensionality

Sources of intensionality

Coding

Expressing properties

Self-reference

Conclusion and open questions

Philosophical significance

The framework

Here I am interested in arithmetical theories. This may be PA or only EA or Q. For the moment being, I also do not commit myself to any specific language or theory, except that it is always assumed that the underlying logic is a standard first-order language, but will make more specific assumptions when needed.

Philosophical significance

Many results in metamathematics have been thought to be philosophically significant, above all the Gödel incompleteness theorems.

Many of the results depend on what a sentence in the language of arithmetic expresses. But there may be different sentences expressing a certain metatheoretic claim, or we may doubt whether a given arithmetical sentence expresses what it is purported to say.

It is well-known that the second incompleteness theorem depends on how provability is expressed. But there are further sources of intensionality.

Examples

The following claims are 'versions' of mathematical results in one of their 'philosophically usable' forms. I do not claim that they are correct; they are certainly not precise.

- No theory extending PA proves its own consistency.
- The sentence stating its own unprovability in Q is independent in Q and it is true.
- The sentence stating its own provability in PA is provable.
- The sentence stating its own Σ_1-truth is refutable.
- The sentence stating its own Π_1-truth is provable.

I am also interested in claims in extensions of the language of arithmetic with a primitive predicate for truth, necessity, knowledge, or still another notion (e.g., Visser–Yablo paradox).

Examples

The following claims are ‘versions’ of mathematical results in one of their ‘philosophically usable’ forms. I do not claim that they are correct; they are certainly not precise.

- No theory extending PA proves its own consistency.
- The sentence stating its own unprovability in Q is independent in Q and it is true.
- The sentence stating its own provability in PA is provable.
- The sentence stating its own Σ_1-truth is refutable.
- The sentence stating its own Π_1-truth is provable.

I am also interested in claims in extensions of the language of arithmetic with a primitive predicate for truth, necessity, knowledge, or still another notion (e.g., Visser–Yablo paradox).

Examples

The following claims are 'versions' of mathematical results in one of their 'philosophically usable' forms. I do not claim that they are correct; they are certainly not precise.

- No theory extending PA proves its own consistency.
- The sentence stating its own unprovability in Q is independent in Q and it is true.
- The sentence stating its own provability in PA is provable.
- The sentence stating its own Σ_1-truth is refutable.
- The sentence stating its own Π_1-truth is provable.

I am also interested in claims in extensions of the language of arithmetic with a primitive predicate for truth, necessity, knowledge, or still another notion (e.g., Visser–Yablo paradox).

Examples

The following claims are 'versions' of mathematical results in one of their 'philosophically usable' forms. I do not claim that they are correct; they are certainly not precise.

- No theory extending PA proves its own consistency.
- The sentence stating its own unprovability in Q is independent in Q and it is true.
- The sentence stating its own provability in PA is provable.
- The sentence stating its own Σ_1-truth is refutable.
- The sentence stating its own Π_1-truth is provable.

I am also interested in claims in extensions of the language of arithmetic with a primitive predicate for truth, necessity, knowledge, or still another notion (e.g., Visser–Yablo paradox).

Examples

The following claims are 'versions' of mathematical results in one of their 'philosophically usable' forms. I do not claim that they are correct; they are certainly not precise.

- No theory extending PA proves its own consistency.
- The sentence stating its own unprovability in Q is independent in Q and it is true.
- The sentence stating its own provability in PA is provable.
- The sentence stating its own Σ_1-truth is refutable.
- The sentence stating its own Π_1-truth is provable.

I am also interested in claims in extensions of the language of arithmetic with a primitive predicate for truth, necessity, knowledge, or still another notion (e.g., Visser–Yablo paradox).

The metatheoretic claims

The metatheoretic claims we are interested are the sentences expressed in our metatheoretic language, e.g.:

- PA is consistent.
- This sentence is provable.
- This sentence is Σ_1-true.

The use of the definite article in ‘the sentence’ is a simplification at best.

Self-reference

I am especially interested in self-referential sentences, sentences that say something about themselves, ascribe to themselves.

> *Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet.* Gödel (1931, p. 175)
>
> *We thus have a sentence in front of us that claims its own unprovability.*

Intensionality

General intensionality

Rough characterization of general intensionality

A metatheoretic claim is intensional iff it depends on which metatheoretic statement (or property) is expressed by a sentence or formula of arithmetic.

By metatheory I mean here our informal metatheory that is not arithmetized.

Every sufficiently strong and sound r.e. arithmetical theory is incomplete: extensional

Logical validity is not decidable: extensional

The sentence stating its own unprovability in Q is independent in Q and it is true: intensional

The sentence stating its own Σ_1-truth via $\mathrm{Tr}_{\Sigma_1}(x)$ in PA with Feferman's (1960) coding is refutable: intensional

Every sufficiently strong and sound r.e. arithmetical theory is incomplete: extensional

Logical validity is not decidable: extensional

The sentence stating its own unprovability in Q is independent in Q and it is true: intensional

The sentence stating its own Σ_1-truth via $\mathrm{Tr}_{\Sigma_1}(x)$ in PA with Feferman's (1960) coding is refutable: intensional

Every sufficiently strong and sound r.e. arithmetical theory is incomplete: extensional

Logical validity is not decidable: extensional

The sentence stating its own unprovability in Q is independent in Q and it is true: intensional

The sentence stating its own Σ_1-truth via $\mathrm{Tr}_{\Sigma_1}(x)$ in PA with Feferman's (1960) coding is refutable: intensional

Every sufficiently strong and sound r.e. arithmetical theory is incomplete: extensional

Logical validity is not decidable: extensional

The sentence stating its own unprovability in Q is independent in Q and it is true: intensional

The sentence stating its own Σ_1-truth via $\mathrm{Tr}_{\Sigma_1}(x)$ in PA with Feferman's (1960) coding is refutable: intensional

Extensionalization

We can make intensional results more specific until it doesn't matter what an arithmetical sentence expresses.

For instance, we can specify a specific coding, a specific provability predicate $\mathrm{Bew}_{\mathsf{PA}}(x)$ and prove $\mathsf{PA} \nvdash \neg\mathrm{Bew}_{\mathsf{PA}}(\ulcorner 0=1 \urcorner)$. This is then an extensional results. But it's philosophically not more usable for the usual purposes than the first incompleteness theorem unless we have some evidence that $\neg\mathrm{Bew}_{\mathsf{PA}}(\ulcorner 0=1 \urcorner)$ expresses the consistency of PA.

Extensionalization

The problems of extensionalization are not only the usual problems of formalization for the following reasons (among others):

1. Arithmetical theories are about numbers, not sentences, formulæ, proofs, sequences of sentences and the like. A Gödel numbering is required.
2. Very often there is no specific sentence to formalize, because the claim is general such as in 'No r.e. theory extending $\mathsf{I}\Sigma_1$ proves its own consistency.'
3. We often aim at using limited resources (bounded quantifiers) and define Σ_n-truth without appealing to the standard model.

Philosophical significance again

The worry is that the philosophical significance of metamathematical results is undermined by intensionality.

For instance, the philosophical significance of the second incompleteness theorem is threatened by the existence of 'provability predicates' (e.g. Rosser provability) for which the second incompleteness theorem fails.

A result may hold for one specific coding scheme, but not for others (e.g. non-monotonic coding schemes).

A result may hold for one specific way of diagonalizing a formula, but not for others.

Philosophical significance again

The logician can react by proving that a result is stable or invariant under variations of relevant parameters (coding, provability or truth predicates).

If this is not possible, an argument is needed why a specific choice is correct, but not others. For instance, we may have to explain why certain coding schemata for which a result fails are irrelevant.

Sources of intensionality

Sources of intensionality

Intensionality has various sources (and consequently extensionalization has to deal with these sources).

1. language (function symbols for all p.r. functions, for exponentiation, or fewer)
2. coding of expressions (sentences)
3. expressing properties such as partial truth, provability etc, expressing consistency etc.
4. self-reference

The sources and their order

Which sources are relevant depends on the claim. For instance, intensionality from self-reference is irrelevant for the second incompleteness theorem (which is not to say that we have proofs of G2 without self-reference).

The sources are ordered: later sources depend on earlier sources: E.g. whether a formula is a provability predicate depends on the coding; whether a sentence says about itself that it's provable depends on which provability predicate is used.

Sources and criteria

Different sources of intensionality are usually addressed in very different ways. Approaches are often very *ad hoc* and specific to one particular claim.

- 'natural' codings
- 'natural' provability predicate
- $\underbrace{S \ldots S}_{n} 0$ as quotation of the expression with code n
- 'canonical' diagonalization

I will now go through the last three sources and assume a fixed language and theory and focus on PA.

Coding

Common practice

Usually in metamathematics people fix some coding and prove their results relative to that coding. Then they may add that their proof goes through for other ‘reasonable’ coding schemata.

The coding has to be reasonable; but it’s hardly made explicit what ‘reasonable’ means. The usual operations on codes should be computable and be provably recursive in the relevant theory. Thus, a coding schema may be reasonable relative to PA, but fail to be reasonable relative to a weaker theory.

Example: G1

For a proof of G1, choose a coding such that for every $\varphi(x)$ there is an n s.t. $n = \ulcorner\varphi(\overline{n})\urcorner$ (‘coding with built-in self-reference’) (Visser, see the Grabmayr–Visser paper).

Relative to that coding, assume that $\mathrm{Bew}(x)$ is a provability predicate.

$$\mathsf{PA} \vdash \overline{n} = \overline{\ulcorner\neg\mathrm{Bew}(\overline{n})\urcorner} \qquad \text{coding}$$

$$\mathsf{PA} \vdash \neg\mathrm{Bew}(\overline{n}) \leftrightarrow \neg\mathrm{Bew}(\overline{\ulcorner\neg\mathrm{Bew}(\overline{n})\urcorner}) \qquad \text{logic}$$

Proceed as usual to show $\mathsf{PA} \nvdash \neg\mathrm{Bew}(\overline{n})$ and $\mathsf{PA} \nvdash \neg\neg\mathrm{Bew}(\overline{n})$.

This proof hardly supports the claim that the sentence stating its own unprovability is independent, because this proof does not work for the usual coding schemata (even if we ignore worries about self-reference).

Example: Truth teller

There is a Σ_1-formula $Tr_{\Sigma_1}(x)$ that is said to be a Σ_1-truth predicate; it provably satisfies all T-sentences for Σ_1-sentences as well as the compositional axioms for such sentences.

Σ_1-truth teller

The Σ_1-truth teller is refutable. That is, the sentence stating its own Σ_1-truth is refutable.

Here we focus on *coding*. We grant that $Tr_{\Sigma_1}(x)$ expresses Σ_1-truth and canonical diagonalization yields self-reference. Moreover, we assume we have function symbols such that for every $\varphi(x)$ there is a closed term t obtained in the canonical way such that $t = \overline{\ulcorner \varphi(t) \urcorner}$.

The canonical truth predicate $\mathrm{Tr}_{\Sigma_1}(x)$ is of the form $\exists y\, \vartheta(y, x)$ with a formula $\vartheta(y, x)$ containing only bounded quantifiers.

THEOREM

Suppose we employ a monotone Gödel coding and $d(\mathrm{Tr}_{\Sigma_1}(x))$ is the canonical diagonal sentence, then $\mathsf{PA} \vdash \neg d(\mathrm{Tr}_{\Sigma_1}(x))$.

assumption

If $\exists v\, \sigma(v)$ is a Σ_1-sentence, that is, if the formula $\sigma(v)$ contains no unbounded quantifier and only the variable v is free in $\sigma(v)$, then $\mathrm{PA} \vdash \forall y\, (\vartheta(y, \overline{\ulcorner \exists v\, \sigma(v) \urcorner}) \rightarrow \exists v < y\, \sigma(v))$ holds.

The truth teller $d(\mathrm{Tr}_{\Sigma_1}(x))$ is of the form $\exists y\, \vartheta(y, t)$ where t is a term denoting this very sentence. Thus, $t = \overline{\ulcorner \exists y\, \vartheta(y, t) \urcorner}$ is true and, hence, PA-provable.

We reason in PA. Suppose $\exists y\, \vartheta(y, t)$. Let y_0 be the smallest witness of $\exists y\, \vartheta(y, t)$. So, (a) $\vartheta(y_0, t)$ and (b) $\forall z < y_0\, \neg\vartheta(z, t)$. Since $t = \overline{\ulcorner \exists y\, \vartheta(y, t) \urcorner}$, the assumption above combined with (a), gives us $\exists z < y_0\, \vartheta(z, t)$. But this contradicts (b). Hence our supposition that $\exists y\, \vartheta(y, t)$ must fail.

assumption

If $\exists v\, \sigma(v)$ is a Σ_1-sentence, that is, if the formula $\sigma(v)$ contains no unbounded quantifier and only the variable v is free in $\sigma(v)$, then $\mathrm{PA} \vdash \forall y\, (\vartheta(y, \overline{\ulcorner \exists v\, \sigma(v) \urcorner}) \rightarrow \exists v < y\, \sigma(v))$ holds.

Thus we have a proof that the Σ_1-truth teller is refutable, but only with the assumption, which follows from a reasonable coding schema. The assumption is technical. Depending on how exactly $\mathrm{Tr}_{\Sigma_1}(x)$ is defined, the assumption follows from

- the assumption that the code of a sentence $\varphi(\overline{n})$ is always greater than n and this property is transmitted through sequences of sentences, or
- the assumption that the code of a variable assignment is always greater than that of any entry and this property is transmitted through sequences – or some such assumption.

Invariance

Very often, people claim that a proofs goes through also for all other ‘reasonable’ coding without offering proofs.

Proving invariance results is made harder because we need to show that proofs go through with the coding changed, but the formulæ expressing provability, Σ_1-truth etc. ‘still the same’. Of course, the formulæ need modifying, but somehow their ‘structure’ should stay the same.

Invariance under changes of the coding

A coding schema is understood as a relative interpretation of a fixed syntax theory into the arithmetical theory (assuming the vocabulary is finite).

The relative interpretations induce translations back-and-forth between different coding schemata.

This allows us to prove the equivalence of all Gödel sentences based on different codings, but built from a canonical provability predicate.

We should also be able to prove the invariance of G2 under changes of the coding (but insisting on canonical provability).

The refutability of the Σ_1-truth teller may fail to be invariant.

Expressing properties

Properties

There may be different ways to express provability in PA, Σ_n-truth, logical truth, etc.

The problem arises when one tries to generalize G2 to all extensions of PA (or some suitable theory), because we need to have a provability predicate for each such extension.

This is the third paragraph of Kreisel's 1953 paper with the notation adapted:

KREISEL'S CRITERION FOR THE EXPRESSION OF PROVABILITY

A formula $\mathrm{Bew}(x)$ is said to express provability in Σ if it satisfies the following condition: for numerals $\overline{n}$, $\mathrm{Bew}(\overline{n})$ can be proved in Σ if and only if the formula with number n can be proved in Σ.

Feferman's notion of intensionality

Feferman 1960, p. 35:

> *In broad terms, the applications of the method [of arithmetization] can be classified as being extensional if essentially only numerically correct definitions are needed, or intensional if the definitions must more fully express the notions involved, so that various of the general properties of these notions can be formally derived.*

Of course, this notion of intensionality has little to do with what philosophers nowadays call intensionality. But see (Carnap 1934, §71).

meaning postulates and similarity

Numeralwise representability, will give us G1 in some intensional form, but not G2.

Moreover, this notion of representability is useless for more complex notions, such as Π_1-truth or the like.

Alternative: 'meaning' postulates:

- Whatever satisfies the Löb derivability conditions expresses provability.
- Whatever provably yields the T-sentences is a truth predicate (for a specific class of sentences).

We may want to insist on ‘natural’ formalizations. But then we probably we do not have a Σ_1-truth predicate in the language of arithmetic.

Example: The Σ_1-truth teller again

We have three Σ_1-truth tellers; one is refutable, one provable, one independent. All truth predicates represent Σ_1-truth numeralwise. CT is the axiomatic 'Tarskian' theory. From (Halbach and Leigh 2021):

THEOREM

- The canonical diagonal sentence of Tr_{Σ_1} is refutable in PA and CT, as long as the coding is monotone.
- Every diagonal sentence of $\mathrm{Bew}_{\mathrm{I}\Sigma_1}$ is provable in PA and CT, where $\mathrm{Bew}_{\mathrm{I}\Sigma_1}$ is canonical. (McGee)
- The canonical diagonal sentence of the primitive truth predicate T in CT is independent of CT.

This leads as to the final source of intensionality…

Self-reference

Ascribing properties to oneself

Let a coding schema and a formula $\varphi(x)$ expressing property P be fixed.

Which sentences say of themselves that they have property P?

Being a diagonal sentence of $\varphi(x)$ is a necessary condition at best. $0 = 0$ is a diagonal sentence of $\mathrm{Tr}_{\Sigma_1}(x)$ and of $\mathrm{Bew}_{\mathsf{PA}}(x)$, but not a truth teller or Henkin sentence.

For canonical provability we do not have to look at the notion of self-reference, because all diagonal sentences behave in the same way by Löb's theorem.

Ascribing properties to oneself

Let a coding schema and a formula $\varphi(x)$ expressing property P be fixed.

Which sentences say of themselves that they have property P?

Being a diagonal sentence of $\varphi(x)$ is a necessary condition at best. $0 = 0$ is a diagonal sentence of $\mathrm{Tr}_{\Sigma_1}(x)$ and of $\mathrm{Bew}_{\mathsf{PA}}(x)$, but not a truth teller or Henkin sentence.

For canonical provability we do not have to look at the notion of self-reference, because all diagonal sentences behave in the same way by Löb's theorem.

The Kreisel–Henkin criterion

KREISEL–HENKIN CRITERION FOR SELF-REFERENCE

Let a formula $\varphi(x)$ expressing a certain property P in Σ be given. Then a formula γ *says about itself that it has property P* iff it is of the form $\varphi(t)$ for some closed term t that has (the code of) $\varphi(t)$ as its value.

If γ says about itself that it has property P (expressed by $\varphi(x)$), then γ is a diagonal sentence of $\varphi(x)$.

Proof: γ must be of the form $\varphi(t)$ where t has $\varphi(t)$ as its value. Since Σ decides all closed equations, we have $\Sigma \vdash t = \overline{\ulcorner\varphi(t)\urcorner}$. From $\Sigma \vdash \varphi(t) \leftrightarrow \varphi(t)$ by substitution of identicals we get $\Sigma \vdash \varphi(t) \leftrightarrow \varphi(\overline{\ulcorner\varphi(t)\urcorner})$.

The Kreisel–Henkin criterion

The criterion is perhaps a sufficient criterion. If it were necessary, we couldn't have self-reference in ZF and in PA (for all codings).

But results may still be unstable under different ways of obtaining self-reference in the sense of the Kreisel–Henkin criterion.

Provable and refutable Henkin sentences

The following is based on Kreisel (1953). We work in a language with sufficiently many function symbols.

THEOREM

Similarly, there is a provability predicate $\mathrm{Bew}_2(x)$ and a term t_2 such that

(i) $\mathrm{Bew}_2(x)$ weakly represents provability in Σ.

(ii) $\Sigma \vdash t_2 = \overline{\ulcorner \mathrm{Bew}_2(t_2) \urcorner}$

(iii) $\Sigma \vdash \neg\mathrm{Bew}_2(t_2)$

(iv) Applying the canonical diagonal procedure to $\mathrm{Bew}_2(x)$ yields a term t with $\Sigma \vdash \mathrm{Bew}_2(t)$

$\mathrm{Bew}_2(t_2)$ is the refutable, $\mathrm{Bew}_2(t)$ the provable Henkin sentence. Both are self-referential in the Kreisel–Henkin sense.

Proof

Fix some predicate $\mathrm{Bew}(x)$ that weakly represents Σ-provability in Σ. By Gödel's diagonal lemma there is a term t_2 such that

$$(1) \qquad \Sigma \vdash t_2 = \overline{\ulcorner t_2 \neq t_2 \wedge \mathrm{Bew}(t_2) \urcorner}$$

Now define $\mathrm{Bew}_2(x)$ as

$$x \neq t_2 \wedge \mathrm{Bew}(x)$$

Clearly $\Sigma \vdash t_2 = \overline{\ulcorner t_2 \neq t_2 \wedge \mathrm{Bew}(t_2) \urcorner}$ and hence (ii) holds by (1). Since

$$t_2 \neq t_2 \wedge \mathrm{Bew}(t_2)$$

is refutable in pure logic (and thus in Σ), $\Sigma \vdash \neg\mathrm{Bew}_2(t_2)$ and (iii) is satisfied.

$\mathrm{Bew}_2(x)$ is a deviant provability predicate, of course.

Are there any natural predicates expressing some metatheoretic property that are sensitive in this way to how self-reference is obtained?

I feel that if we are interested in *the* sentence ascribing to itself the property of being provable via $\mathrm{Bew}_2(x)$, we should look at the canonical diagonal sentence $\mathrm{Bew}_2(t)$. $\mathrm{Bew}_2(t_2)$ is only 'accidentally' self referential.

Example: Apply canonical diagonalization to $x = x$ to obtain a term s with $\Sigma \vdash s = \overline{\ulcorner s = s \urcorner}$. Now consider the property of being identical with s, expressed by $x = s$. Apply canonical diagonalization to $x = s$ to obtain a term t with $t = \overline{\ulcorner t = s \urcorner}$.

Both $s = s$ and $t = s$ say about themselves that they are identical with s in the sense of the Kreisel–Henkin criterion; but $s = s$ does so only 'accidentally'.

Balthasar and Lingyuan are now trying to make the notion of 'uniform self-reference' precise. That is, we look at sentences that are obtained using a uniform, non-accidental method for generating self-reference.

Conclusion and open questions

Conclusion

Problems for the philosophical significance of intensional metamathematical results may not only arise through the way provability, truth etc. are expressed, but also through coding and the way self-reference is obtained.

The three sources of self-reference interact. Being more strict on say coding, may give us invariance in another dimension. Insisting on a specific provability predicate may give us invariance under different ways of obtaining self-reference.

By extensionalizing results and proving invariance results we can make metatheoretic results more philosophically significant.

Although it is often said that e.g. goes through for all other reasonable codings, we should make clear what reasonable codings are.

Conclusion

Problems for the philosophical significance of intensional metamathematical results may not only arise through the way provability, truth etc. are expressed, but also through coding and the way self-reference is obtained.

The three sources of self-reference interact. Being more strict on say coding, may give us invariance in another dimension. Insisting on a specific provability predicate may give us invariance under different ways of obtaining self-reference.

By extensionalizing results and proving invariance results we can make metatheoretic results more philosophically significant.

Although it is often said that e.g. goes through for all other reasonable codings, we should make clear what reasonable codings are.

Conclusion

Problems for the philosophical significance of intensional metamathematical results may not only arise through the way provability, truth etc. are expressed, but also through coding and the way self-reference is obtained.

The three sources of self-reference interact. Being more strict on say coding, may give us invariance in another dimension. Insisting on a specific provability predicate may give us invariance under different ways of obtaining self-reference.

By extensionalizing results and proving invariance results we can make metatheoretic results more philosophically significant.

Although it is often said that e.g. goes through for all other reasonable codings, we should make clear what reasonable codings are.

Conclusion

Problems for the philosophical significance of intensional metamathematical results may not only arise through the way provability, truth etc. are expressed, but also through coding and the way self-reference is obtained.

The three sources of self-reference interact. Being more strict on say coding, may give us invariance in another dimension. Insisting on a specific provability predicate may give us invariance under different ways of obtaining self-reference.

By extensionalizing results and proving invariance results we can make metatheoretic results more philosophically significant.

Although it is often said that e.g. goes through for all other reasonable codings, we should make clear what reasonable codings are.

Open (and closed) questions I

Are there better criteria than the Kreisel–Henkin criterion for self-reference?

How can we rule out 'accidental' self-reference and should we?

Can all the results (like refutable Henkin sentences) always be obtained by tweaking the representing formula while canonical diagonalisation is retained? (Visser showed there are refutable Henkin sentences obtained via canonical diagonalization.)

Open (and closed) questions I

Are there better criteria than the Kreisel–Henkin criterion for self-reference?

How can we rule out ‘accidental’ self-reference and should we?

Can all the results (like refutable Henkin sentences) always be obtained by tweaking the representing formula while canonical diagonalisation is retained? (Visser showed there are refutable Henkin sentences obtained via canonical diagonalization.)

Open (and closed) questions I

Are there better criteria than the Kreisel–Henkin criterion for self-reference?

How can we rule out ‘accidental’ self-reference and should we?

Can all the results (like refutable Henkin sentences) always be obtained by tweaking the representing formula while canonical diagonalisation is retained? (Visser showed there are refutable Henkin sentences obtained via canonical diagonalization.)

Open (and closed) questions

Is there any 'natural' predicate with two fixed points that both satisfy the Kreisel–Henkin criterion but differ in their properties?

What happens with other properties such as Rosser provability? This question has been answered by Kurahashi (2014).

Can we get provable or independent Σ_1-truth tellers using canonical truth predicates, but non-monotonic codings?

Ar there more uniform criteria for expressing a property?

Open (and closed) questions

Is there any 'natural' predicate with two fixed points that both satisfy the Kreisel–Henkin criterion but differ in their properties?

What happens with other properties such as Rosser provability? This question has been answered by Kurahashi (2014).

Can we get provable or independent Σ_1-truth tellers using canonical truth predicates, but non-monotonic codings?

Ar there more uniform criteria for expressing a property?

Open (and closed) questions

Is there any 'natural' predicate with two fixed points that both satisfy the Kreisel–Henkin criterion but differ in their properties?

What happens with other properties such as Rosser provability? This question has been answered by Kurahashi (2014).

Can we get provable or independent Σ_1-truth tellers using canonical truth predicates, but non-monotonic codings?

Ar there more uniform criteria for expressing a property?

Open (and closed) questions

Is there any ‘natural’ predicate with two fixed points that both satisfy the Kreisel–Henkin criterion but differ in their properties?

What happens with other properties such as Rosser provability? This question has been answered by Kurahashi (2014).

Can we get provable or independent Σ_1-truth tellers using canonical truth predicates, but non-monotonic codings?

Ar there more uniform criteria for expressing a property?

References

Rudolf Carnap. *Logische Syntax der Sprache*. Springer, Wien, 1934.

Solomon Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–91, 1960.

Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik*, 38:173–198, 1931.

Volker Halbach and Graham Leigh. *The Road to Paradox: A Guide to Syntax, Truth, and Modality*. Cambridge University Press, 2021. to appear.

Georg Kreisel. On a problem of Henkin's. *Indagationes Mathematicae*, 15:405–406, 1953.

Taishi Kurahashi. Henkin sentences and local reflection principles for Rosser provability. to appear, 2014.

Peter Milne. On Gödel Sentences and What They Say. *Philosophia Mathematica*, 15:193–226, 2007.