A Closer Look at Self-Reference in Arithmetic: Language, Naming and Coding

There are essentially two ways an arithmetical sentence can be taken to (directly) refer to itself, namely, in virtue of containing a term denoting (the code of) itself, or by means of quantification (see (Leitgeb, 2002), (Picollo, 2018)). In this talk we are only concerned with the first kind of self-reference, also called "self-reference by mention" or in short "m-self-reference" (Picollo, 2018, p. 581). Halbach and Visser (2014a) trace this notion of self-reference back to the famous exchange of Kreisel and Henkin regarding Henkin's question as to whether sentences that state their own PA-provability are provable in PA. Heck (2007), for instance, considers m-self-reference the only legitimate way to formalise truly self-referential reasoning in arithmetic.

The primary concern of this talk is to examine the conditions under which m-self-reference is attainable in arithmetic. There are two well-known routes to achieve this. Let $\mathcal{L}$ be the arithmetical language which has 0, S, $+$ and $\times$ as its non-logical vocabulary.

The first route consists in enriching the arithmetical language. That is, function symbols for certain primitive recursive functions needed to formally represent the (canonical) diagonal operator are added to $\mathcal{L}$, yielding $\mathcal{L}^+$. Moreover, the equations that primitive recursively define the relevant functions are added to a given arithmetical theory $T \supseteq \mathsf{R}$,[1] yielding T^+. Let ξ be a standard Gödel numbering of $\mathcal{L}^+$ and let $\underline{n}$ denote the standard numeral of n. The Strong Diagonal Lemma then provides the existence of m-self-referential sentences:

Lemma 1 (Strong Diagonal Lemma) *Let A be a $\mathcal{L}^+$-formula with exactly x as a free variable. Then there exists a closed term t in $\mathcal{L}^+$ such that $T^+ \vdash t = \underline{\xi(A(t))}$.*

The second route is based on *self-referential* Gödel numberings. A numbering ξ is called self-referential if, for any formula $A(x)$, we can find a number n such that $n = \xi(A(\underline{n}))$. Self-referential numberings thus immediately provide m-self-referential sentences, without the need of extending the language (as well as the theory). The idea of self-referential Gödel numberings can be traced back to (Kripke, 1975, footnote 6). Constructions of self-referential numberings are given in (Visser, 1989), (Heck, 2007), (Kripke, 2020) and in the Appendix of (Halbach and Visser, 2014b).

The second route is typically considered to be contrived and unsatisfactory since the self-referential numberings existing in the literature to date do not satisfy desirable properties. One such property is *monotonicity*, which requires the Gödel number of an expression to be larger than the Gödel number of a sub-expression. Halbach (2018), for instance, does not consider non-monotonic numberings as adequate choices of Gödel numberings. Indeed, the study of self-reference is commonly based on numberings which are required to be monotonic. It is, thus, widely believed that for adequate choices of Gödel numberings m-self-reference is not attainable in arithmetic formulated in $\mathcal{L}$ (see (Heck, 2007)). Accordingly, the natural setting to formalise self-referential reasoning is typically based on the first approach, by increasing the (term-)expressiveness of the language $\mathcal{L}$.

The aim of the present talk is to further investigate this trade-off between the expressiveness of the language and the naturalness of the underlying coding. In particular, we will show that the received view regarding the unattainability of m-self-reference in $\mathcal{L}$ is committed to much stronger assumptions on the coding apparatus than usually assumed in the literature.

[1] R denotes the Tarski-Mostowski-Robinson theory introduced in (Tarski et al., 1953, p. 53).

In this talk, we will show that when employing standard numerals, self-referential Gödel numberings indeed cannot be monotonic. Thus, by Halbach's standards, self-referential Gödel numberings would be disqualified as adequate formalisation choices.

This is however not the end of the story. We will show that we can produce effective self-referential monotonic numberings by basing the definition of *self-referential* on efficient instead of standard numerals. Efficient numerals form another kind of naming device that corresponds to a dyadic notation system, frequently used in the literature on weak arithmetical systems. That is, for any formula $A(x)$, we can find a number n such that $n = \xi(A(\overline{n}))$, where $\overline{n}$ denotes the efficient numeral of n. Setting $t := \overline{n}$ hence yields $\mathsf{R} \vdash t = \underline{\xi(A(t))}$. The adequacy constraints on numberings have thus to be strictly more restrictive than effectiveness and monotonicity in order to rule out self-referential numberings and, in particular, m-self-referentiality in $\mathcal{L}$.

We will further introduce a strengthened notion of monotonicity put forward by Halbach (2018), which captures the idea that (efficient) numerals are arithmetical proxies for quotations. A monotonic coding is called strongly monotonic if the code of the Gödel numeral of an expression is larger than the code of the expression itself. This constraint on numberings is sufficiently restrictive to exclude self-referential numberings for any numeral system. However, we will show that even strong monotonicity is not restrictive enough to exclude m-self-referentiality, i.e., strong diagonalisation. In fact, we will construct an effective and strongly monotonic numbering which gives rise to the Strong Diagonal Lemma for $\mathcal{L}$, thus providing m-self-referential sentences formulated in $\mathcal{L}$.

We moreover introduce computational constraints which are more restrictive than effectiveness and which may serve as additional adequacy constraints for numberings. A Gödel numbering is called $\mathfrak{El}$-adequate if it represents a large portion of syntactic relations and operations by *elementary* relations and operations on ω. We show that the numberings constructed in this talk are $\mathfrak{El}$-adequate in this sense. Hence, strong monotonicity and $\mathfrak{El}$-adequacy taken together, are once again not restrictive enough to exclude m-self-referentiality in $\mathcal{L}$.

The obtained results suggest the following disjunctive conclusion: when formalising m-self-reference in arithmetic, the adequacy constraints on reasonable numberings are more restrictive than widely assumed, or m-self-reference can be adequately formalised in a less expressive language than usually believed. More specifically, we conclude that either the constraints on reasonable numberings are more restrictive than $\mathfrak{El}$-adequacy and strong monotonicity taken together, or m-self-reference is already attainable in $\mathcal{L}$.

Time permitting, we will show how these results bear on the study of axiomatic truth theories. In particular, we show that the constraints of $\mathfrak{El}$-adequacy and (strong) monotonicity taken together are not sufficient to determine the consistency of certain type-free truth theories. Thus, the formalisation of certain informal principles of truth in an arithmetical setting is highly sensitive to the underlying formalisation choices. These results raise doubts as to what extent such axiomatic theories can be taken to faithfully reflect informal reasoning regarding the underlying principles of truth.[2]

References

Grabmayr, B. and Visser, A. (2020). Self-reference upfront: A study of self-referential Gödel numberings. *Review of Symbolic Logic*. Forthcoming.

Halbach, V. (2018). Personal communication, March 20.

[2]For more details the reader is referred to (Grabmayr and Visser, 2020).

Halbach, V. and Visser, A. (2014a). Self-reference in arithmetic I. *Review of Symbolic Logic*, 7(4):671–691.

Halbach, V. and Visser, A. (2014b). Self-reference in arithmetic II. *Review of Symbolic Logic*, 7(4):692–712.

Heck, R. K. (2007). Self-reference and the languages of arithmetic. *Philosophia Mathematica*, 15(1):1–29.

Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716.

Kripke, S. A. (2020). Gödel's theorem and direct self-reference. arXiv:2010.11979, available at `https://arxiv.org/abs/2010.11979`.

Leitgeb, H. (2002). What is a self-referential sentence? *Logique and Analyse*, 177(178):3–14.

Picollo, L. (2018). Reference in arithmetic. *Review of Symbolic Logic*, 11(3):573–603.

Tarski, A., Mostowski, A., and Robinson, R. (1953). *Undecidable Theories*. Studies in logic and the foundations of mathematics. North-Holland, Amsterdam.

Visser, A. (1989). Semantics and the liar paradox. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic*, volume IV, pages 617–706. Reidel, Dordrecht.