CANONICAL INCOMPLETENESS

Abstract. In this note we explain to logicians and philosophers how the perspective of categorical logic provides a *canonical* way of understanding why Gödel's incompleteness results are true. This note is currently under construction!

1. Introduction

Gödel proved his famous incompleteness results in [4]. The first incompleteness theorem, that PA contains a true but unprovable sentence, is an extensional statement, whose truth can be grasped clearly and without doubt. However, the second incompleteness theorem, which states that PA cannot prove its own consistency, is an *intensional* statement, whose truth depends on much more subtle details of the arithmetisation process, i.e. how we recognise PA internally in PA itself.

As the classical work by Feferman [2] indicates, externally equivalent but internally different ways of enumerating axioms of PA may in fact *falsify* Gödel's second incompleteness theorem: There exists a "non-standard" enumeration, under which PA can prove its own consistency. Further works also imply that different choices of coding schemes, provability predicates, or internalisation of the consistency statement may all be exploited to invalidate Gödel's second incompleteness theorem; see [6, 7].

From a mathematical point of view, Gödel's first incompleteness theorem being *extensional* is due to the fact that it is a statement about the copy of PA constructed in the meta-theory we choose, say ZF. However, the second incompleteness theorem is a statement about a copy of PA *inside* PA, where we treat PA as a new meta-theory that we work within. The crucial point is that PA as a meta-theory is a much weaker theory than the original meta-theory ZF, such that equivalent formulations of PA in ZF might fail to be equivalent anymore *within* PA. Thus, different ways of formalising PA within PA may lead to non-equivalent logical theories, thus to non-equivalent result concerning the corresponding consistency statement.

Thus, in essense, the heart of the problem related to the intensionality aspect of the second incompleteness theorem is the following: *How do we*

Date: July 3, 2024.

Key words and phrases. Self-reference, Arithmetic, Intensionality, Categorical Logic.

identify the correct copy of PA *in weaker meta-theories?* This note attempts to give a definitive answer to this problem, through the lense of *categorical logic*.

The central reason why category theory might be helpful in such situations is that, the way category theory defines concepts is *universal*, which can naturally be applied to an arbitrary setting. For instance, the way category theory defines product is through a *universal property*. The familiar set-theoretic construction of the Cartesian product of two sets can be viewed as a proof that product exists in the category **Set** of all sets. However, the crucial point of the categorical definition of product is that we can interprete it in different categories. For instance, if we interprete the universal property of product in a *poset*, then we get the notion of *conjunction*, or *meet*.

In a nut shell, the solution provided by categorical logic to the intensionality problem of recognising the correct internal copy of PA is to *give* PA *a universal property*. Thus, we would have a *unique* (up to equivalence) choice of the internal version of PA in PA, or in any other meta-theories we find ourselves in. We will also show that, under this identification, Gödel's second incompleteness theorem is indeed true.

In fact, many of the mathematical contents in this note were already there in the 1970s, explained by Andrè Joyal in a serious lectures, but without publication. Since then, there are various papers [12, 13] explaining the technical aspect of Joyal's original work. This note however makes a philosophical effort in trying to bridge the gap between the technical work involved and the philosophical concerns of logicians and philosophers via well-motivated narratives, to achieve a canonical understanding of Gödel's incompleteness results.

2. Logical Theories as Categories

In this note we consider *coherent logic* and *classical logic*. Coherent theories are first-order theories where all the formulas are constructed out of $\top, \wedge, \bot, \vee, \exists, =$, which will be referred to as *coherent formulas*. The inferences and axioms in coherent theories are formulated in sequent style,

$$\frac{\varphi_0 \vdash \psi_0 \qquad \cdots \qquad \varphi_n \vdash \psi_n}{\varphi \vdash \psi}$$

where all the formulas involved are coherent. For instance, the rules for $\wedge$ and $\top$ are given by three axioms

$$\frac{}{\varphi \vdash \top} \quad \frac{}{\varphi \wedge \psi \vdash \varphi} \quad \frac{}{\varphi \wedge \psi \vdash \psi}$$

and one rule,

$$\frac{\chi \vdash \varphi \qquad \chi \vdash \psi}{\chi \vdash \varphi \wedge \psi}$$

Similarly for other logic connectives. We refer the readers to [10, D1.3] for more detailed introduction on coherent logic.

Classical logic is the full first-order logic with negation $\neg$, and it is *conservative* over coherent logic: For any coherent theory $\mathbb{T}$ and any coherent formulas φ, ψ, if $\varphi \longrightarrow \psi$ is derivable from $\mathbb{T}$ in classical logic, then $\varphi \vdash \psi$ is also derivable from $\mathbb{T}$ in coherent logic.

For our purpose, we will mainly focus on coherent and classical theories of *arithmetic*. The typical classical theory of arithmetic is Peano arithmetic PA. If we restrict ourselves to coherent logic, we can also formulate a coherent theory of arithmetic CA, which can be identified as the theory of sequents between Σ_1-formulas in $I\Sigma_1$. For the details of CA, we refer the readers to [14].

Categorical logic in some sense is a methodology, which offers an alternative understanding of logical theories and their semantics. In this section we will describe the first important thesis behind categorical logic:

Thesis 1. *Coherent (resp. classical) theories can be identified as certain types of categories, called* coherent (resp. Boolean) *categories.*[1]

2.1. **From Theories to Categories.** As a first approximation, coherent and Boolean categories are categories with some additional categorical properties. To *identify* coherent and classical theories as these types of categories, we have the following correspondence:

- Any coherent (resp. classical) theory $\mathbb{T}$ induces a coherent (resp. Boolean) category $\mathcal{C}[\mathbb{T}]$, which fully records both the proof- and model-theoretic properties of $\mathbb{T}$.
- Any coherent (resp. Boolean) category $\mathcal{C}$ also has a corresponding coherent (resp. classical) theory $\mathbb{T}_{\mathcal{C}}$, such that $\mathcal{C} \simeq \mathcal{C}[\mathbb{T}_{\mathcal{C}}]$.

Since we mostly care about logical theories in this note, we mainly focus on the first point, and discuss in more detail what the category $\mathcal{C}[\mathbb{T}]$ look like, and in which sense could we say that $\mathcal{C}[\mathbb{T}]$ fully records the syntactic and semantic information of $\mathbb{T}$.

Let $\mathbb{T}$ be a coherent (resp. classical) theory. The construction of $\mathcal{C}[\mathbb{T}]$ is a generalisation of the Lindenbaum-Tarski algebra construction of a propositional theory:

- Objects are *formulas* $\varphi(\overline{x})$ in $\mathbb{T}$ up to α-equivalence;

[1] We will not explain in full detail what a coherent or Boolean category is. We refer the interested readers to [10, A1.4].

- Morphisms $\theta : \varphi(\overline{x}) \longrightarrow \psi(\overline{y})$ are $\mathbb{T}$-*provably functional formulas* $\theta(\overline{x}, \overline{y})$ between them up to $\mathbb{T}$-provable equivalence.[2]

$\mathcal{C}[\mathbb{T}]$ is usually denoted as the *syntactic category* of $\mathbb{T}$.

2.2. **Proof-Theoretic Aspect of the Syntactic Category.** The way $\mathcal{C}[\mathbb{T}]$ records the proof-theoretic properties of $\mathbb{T}$ depends on a categorical notion of *monomorphisms.* A monomorphism is an appropriate generalisation of *injective functions* between sets but in an arbitrary category.

Let us use $(\overline{x}) = (x_1, \cdots, x_n)$ to denote the following formula

$$(\overline{x}) := \bigwedge_{i=1}^{n} x_i = x_i.$$

$(\overline{x})$ can thus we viewed as *truth* with free variables $\overline{x}$. Then for any formula $\varphi(\overline{x})$, there is a natural monomorphism

$$\varphi(\overline{x}) \rightarrowtail (\overline{x}) \in \mathcal{C}[\mathbb{T}],$$

which can be intuitively read as $\varphi(\overline{x})$ implies truth. All the monomorphisms mapping into $(\overline{x})$ naturally forms a *preorder*, with $\varphi(\overline{x}) \leq \psi(\overline{x})$ in this preorder iff there exists $\varphi(\overline{x}) \rightarrowtail \psi(\overline{x})$ making the following diagram commute,

$$\begin{array}{ccc} \varphi(\overline{x}) & \rightarrowtail & \psi(\overline{x}) \\ & \searrow \quad \swarrow & \\ & (\overline{x}) & \end{array}$$

We will denote the postal reflection of this preorder as $\mathrm{Sub}(\overline{x})$, and call it the *subobject poset* of $\mathrm{Sub}(\overline{x})$. The upshot is the following result:

Fact 2.1. *For any formulas $\varphi(\overline{x})$, $\psi(\overline{x})$, we have*

$$\mathbb{T} \vdash (\varphi(\overline{x}) \vdash \psi(\overline{x})) \iff \varphi(\overline{x}) \leq \psi(\overline{x}) \in \mathrm{Sub}(\overline{x}).$$

Thus, the provability of sequents in $\mathbb{T}$ are completely recorded in subobject posets of the form $\mathrm{Sub}(\overline{x})$, for all lists of variables $(\overline{x})$.

Remark 2.2. In fact, if $\mathbb{T}$ is coherent (resp. classical), then $\mathrm{Sub}(\overline{x})$ will be a *distributive (resp. Boolean)* lattice. The dual stone space of $\mathrm{Sub}(\overline{x})$ is exactly the *space of n-types* in model theory.

[2] Notice that the arrow in "$\theta : \varphi(\overline{x}) \longrightarrow \psi(\overline{y})$" does not signify implication between formulas, but a morphism from $\varphi(\overline{x})$ to $\psi(\overline{y})$ in $\mathcal{C}[\mathbb{T}]$.

2.3. **Model-Theoretic Aspect of the Syntactic Category.** To formulate how $\mathcal{C}[\mathbb{T}]$ also records the model-theoretic information, it is better to discuss a little deeper what are the additional structures required for a category to be coherent, or Boolean. As a second approximation, intuitively a coherent (resp. Boolean) category can be thought of as a category which *supports the interpretation of all the logical connectives in coherent (resp. classical) logic*. For instance, the typical example of a coherent, in fact Boolean, category is the category **Set** of all sets. The coherent or Boolean categorical structure on **Set** is tantamount to the Tarskian truth definition of first-order logic formulas.

From this perspective, we have the following result, which is usually called the *functorial semantics* in categorical logic:

Fact 2.3. *A model M of $\mathbb{T}$ can be equivalently viewed as a* coherent functor

$$M : \mathcal{C}[\mathbb{T}] \longrightarrow \mathbf{Set}.$$

Notice that a functor $M : \mathcal{C}[\mathbb{T}] \longrightarrow \mathbf{Set}$ is an assignment that takes any object in $\mathcal{C}[\mathbb{T}]$ to a set, and takes any morphism to a function. For any model M, the induced functor M should be viewed as assigning any formula $\varphi(\overline{x})$ the set in M *defined by* $\varphi(\overline{x})$:

$$M(\varphi(\overline{x})) := \{ (a_1, \cdots, a_n) \in M^n \mid M, (a_1, \cdots, a_n) \vDash \varphi(\overline{x}) \}.$$

On morphisms, if θ is $\mathbb{T}$-provably functional, then it evidently defines a functional relation on any model M, and thus we get a proper function.

From the categorical perspective, a coherent functor is a functor that preserves the coherent structure on the two coherent categories. According to our intuitive description of the coherent structure in a category as above, this means that the interpretation of formulas $\varphi(\overline{x})$ by the functor M should respect the Tarskian truth definition.

Remark 2.4. One of the important consequences of accepting Thesis 1 is that it provides a nice way of understanding logical theories without worrying about different syntactic choices. One difficulty for recognising a robust and canonical copy of PA with traditional methods of arithmetisation is that there are too many subtle choices we can tweak in the formalisation process of the syntax. The categorical logic approach provides a solution to this problem: As along as different formalisations result in equivalent categories, we would consider them describing the same logical theory.

3. Categories as Metalogic

The second perspective of categorical logic is that, roughly speaking, different categories can be viewed as providing *different meta-theories*, by consider constructions *internally* in a category. As also mentioned in the introduction, most familiar set-theoretic constructions, like products, sums, power sets, function sets, etc., can indeed be expressed as *universal properties*. In particular, it makes sense to ask whether an *arbitrary* category, other than **Set**, supports these constructions. Say if a proof only involves construction of certain types, and these constructions exist in another category $\mathcal{C}$, then the proof could be *internalised* in $\mathcal{C}$. This method of internalisation proves to be very successful [11], and the internalisation process is usually referred to as the working in the *internal logic* of a category.

In categorical logic, there is a typical class of categories, whose internal logic can formalise most of the (constructive) mathematics. They are called *Grothendieck toposes*. Intuitively, they are a class of categories whose internal logic supports all the constructions valid in IZF.

Coherent and *Boolean* categories can be seen as *predicative* versions of Grothendieck toposes, where only general power sets and function sets are not supported. However, for the coherent and Boolean categories $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$, they do supports the non-predicative axiom of infinity, i.e. natural numbers exists.

Again, in the context of categorical logic, the correct interpretation of natural numbers is through a universal property, which allows one to identify in different contexts when an object should be identified as the internal copy of natural numbers. Formally, it is called a *parameterised natural numbers object*, or *PNNO*, in a category. A PNNO expresses a universal property which specifies that this object satisfies the induction principle of natural numbers w.r.t. other objects in this category, and as any other universal property, it characterises such an object uniquely.

Remark 3.1. One way to make this more precise is to consider categorical semantics of *type theories*. Any coherent or Boolean category with PNNO can serve as a *model* of a predicative type system with product types, sum types, quotient types, and natural numbers type. See [12] for a version of such a type theory. The reason here we consider type theory rather than set theory as foundation is that type theory more closely reflects the categorical structures, so that the categorical models of type theories are much easier to formulate; see [8, 9]. Also, through the works [1, 3], one can also translate various versions of set theories in type theories, so there is no essential difference in viewing type theories as foundations.

Fact 3.2. *$\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ will be a coherent, resp. Boolean, category with a PNNO.*

Hence, the meta-theories provided by the internal logics of $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ are versions of predicative foundations with natural numbers. The difference between coherent and Boolean categories in this respect is that the meta-theory provided by the latter is *classical*, while the former is not.

In fact, as we will see in Section 4, CA and PA furthermore have corresponding universal properties w.r.t. the types of categories in Fact 3.2, which, as we've mentioned, is the key towards identifying the unique copy of CA in PA in different context.

The reason we care about the internal logic of $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ is that, developing mathematics in CA or PA in the traditional sense of logicians through the process of *arithmetisation*, is exactly equivalent to doing mathematics in the internal logic of $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$:

Thesis 2. *The arithmetisation process of developing a piece of mathematics in* CA *and* PA *is equivalent to doing mathematics in the internal logic of* $\mathcal{C}[\mathsf{CA}]$ *and* $\mathcal{C}[\mathsf{PA}]$.

Let us discuss in more detail why Thesis 2 is an empirical fact. Recall that we use (x) to denote the formula $x = x$ as an object in $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$. This formula is indeed the PNNO in $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[pa]$.[3] The usual process of arithmetisation of CA or PA in themselves starts with constructing a formula $\mathrm{Fml}(x)$ that enumerates the set of formulas of CA or PA. From the internal perspective, this is exactly a *subset* of natural numbers,

$$\mathrm{Fml}(x) \rightarrowtail (x),$$

which signifies which numbers could be thought of as formulas. The fact that we have definable function that takes the conjunction of two formulas exactly means that, by definition of $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ given in Section 2.1, we can constructing a *map* in $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$ that computes the conjunction of two formulas[4]

$$\dot{\wedge} : \mathrm{Fml}(x) \times \mathrm{Fml}(y) \longrightarrow \mathrm{Fml}(z),$$

where now the product is taken internally in $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$. Similarly for all the other arithmetisation process.

One important insight of Gödel's proof is that, CA and PA have enough expressive power to formalise themselves. Via categorical logic, we get a clear understanding of why this is true from a structural perspective: Under Thesis 1, developing an arithmetic theory internally in a category $\mathcal{C}$ can be viewed as constructing a coherent or Boolean category with a PNNO in the internal logic. However, the notion of a coherent or Boolean categories with PNNO are *algebraic*, and thus can be expressed in a predicative setting. In fact, they can be expressed in much weaker context, which is

[3]One way to under stand this is that (x) defines the set of natural numbers in the standard model.

[4]Here we choose different variables for three occurrences of the formula Fml simply for book keeping reasons, and can be safely ignored. In particular, they are the same object in $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$ since by definition we identify objects up to α-equivalence.

also consistent with the knowledge in logic that PA can be formalised in much weaker logics.

Thus, under Thesis 1 and Thesis 2, the key question to understand the intensional aspect of Gödel's second incompleteness theorem is the following:

Question 3.3. Which internal coherent or Boolean category with a PNNO should be identified as the internal copy of the theory CA or PA?

Notice that at this point, we no longer distinguish CA and PA with their syntactic categories. We provide an answer to this in the next section.

4. Universal Properties of Theories of Arithmetic

As mentioned in Section 1, the key to answer Question 3.3 is to find *universal properties* of CA and PA, so that we can identify in weaker metatheories which is the correct version of CA and PA. We will throw out the punchline directly:

Thesis 3. CA *and* PA *should be identified as the* initial *coherent or Boolean category with a PNNO.*

Let us first provide more detailed information of what Thesis 3 actually means. Suppose we are working in a category $\mathcal{C}$ whose internal logic can define the notion of coherent and Boolean categories with a PNNO. Then in fact, there will be *categories* $\mathsf{CohN}_{\mathcal{C}}$ and $\mathsf{BoolN}_{\mathcal{C}}$ of *internal* coherent and Boolean categories with a PNNO. The morphisms in these categories are internal coherent functors in $\mathcal{C}$ that preserves the PNNO.

For instance, if we work in **Set**, then $\mathsf{CohN}_{\mathbf{Set}}$ and $\mathsf{BoolN}_{\mathbf{Set}}$ will just be the category of coherent and Booleans categories with a PNNO in the usual sense, and we simply abbreviate them as CohN and BoolN. If we work in $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$, then these categories will be the *definable* coherent or Boolean categories with PNNO, with *definable* coherent functors preserving PNNO between them. What Thesis 3 suggests is that, internally in $\mathcal{C}$, the copy of CA and PA should be identified as the *initial object* in the category $\mathsf{CohN}_{\mathcal{C}}$ and $\mathsf{BoolN}_{\mathcal{C}}$, respectively.

Let us first recall what is an initial object in a category. An initial object in a category $\mathcal{D}$ is an object $\varnothing$, such that for *any* object A in $\mathcal{D}$, there is a *unique* arrow in $\mathcal{D}$ from $\varnothing$ to A. For instance, the initial object in **Set** is the empty set. One can also see very easily that the initial objects in $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ is the falsum sentence $\bot$. Just like any other universal property, the initial object in any category is also *unique*, if exists.

Our first empirical justification of Thesis 3 is the following theorem:

Theorem 4.1. $\mathcal{C}[\mathsf{CA}]$ *and* $\mathcal{C}[\mathsf{PA}]$ *are the initial objects of* CohN *and* BoolN*, respectively.*

Proof. See [14]. □

Remark 4.2. One might be puzzled by why Fact 2.1 and Theorem 4.1 can both be true. As Fact 2.1 indicates, coherent functors from $\mathcal{C}[\mathsf{CA}]$ or $\mathcal{C}[\mathsf{PA}]$ to **Set** should be identified as *models* of CA and PA, which by no means are unique, since there exists non-standard models of arithmetic. Thus, this seems to contradict with initiality of $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$. However, notice that in the definition of CohN and BoolN, we also require the functors to *preserve the PNNO*. Given the additional condition, there is then only a unique functor from $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$ to **Set**, viz. the one interpreting the formula (x), which is the PNNO in $\mathcal{C}[\mathsf{CA}]$ and $\mathcal{C}[\mathsf{PA}]$, as the actual set of natural numbers $\mathbb{N}$ in **Set**: These are exactly the standard models of CA and PA.

Philosophically, the initiality result in Theorem 4.1 intuitively means that, CA and PA are in some sense the *minimal* context within which we can reason about natural numbers in a coherent or Boolean category setting. This coheres well with our understanding of the nature of CA and PA, which should result in acceptance of Thesis 3.

Some logicians and philosophers might also argue that PA is by no means a *minimal* classical theory that can reason about natural numbers, since we have weaker arithmetic theories, like $I\Sigma_n$ or even weaker theories of bounded arithmetic. However, this divergence of understanding is caused by different conception of what the object of natural numbers should be. If we take the concept of natural numbers as in the categorical definition of PNNO, then it is an object which has the induction principle w.r.t. other objects. It is precisely in this sense that we claim PA to be the minimal classical theory that can reason about natural numbers, and this is supported exactly by the initiality result in Theorem 4.1.

Thus, Thesis 3 provides an answer to Question 3.3: It characterises CA and PA *uniquely* in *any* other category $\mathcal{C}$.

5. Gödel's Second Incompleteness and Other Intensional Results

Now that we have cleared the conceptual issue related to the intensionality aspect of arithmetisation of CA and PA within themselves, the remaining are technical works to prove Gödel's second incompleteness theorem, or any other intensional meta-logic result in this setting.

As mentioned in Section 1, the main categorical tools to prove such results were already there in the 1970s. We simply record the result here with recent references:

Theorem 5.1. CA *or* PA *cannot prove their own consistency, i.e. they cannot prove the consistency of the initial coherent or Boolean category with PNNO internally in* $\mathcal{C}[\mathsf{CA}]$ *or* $\mathcal{C}[\mathsf{PA}]$.

Proof. Following a similar argument as in [13]. □

Let us also return to the classical result of Feferman [2]. From the perspective of this note, the non-standard enumeration constructed by Feferman, although can be shown to be externally equivalent to the standard enumeration, will indeed *not* be equivalent to the standard enumeration of PA internally in $\mathcal{C}[\mathsf{PA}]$, since it is a much weaker meta-theory compared to the one provided by **Set**. In other words, the logical theory constructed by the non-standard enumeration, result in an internal category which is *not* equivalent to the internal copy of PA, viz. the internal initial Boolean category with PNNO. The suprising fact here is that, by simply changing the enumeration, we can in fact obtain an internal theory which indeed $\mathcal{C}[\mathsf{PA}]$ can show its consistency.

One should be confident that *all* the other meta-logical result which has an intensional aspect can be shown to hold under Thesis 3. We give another example whose proof also already exists in the literature:

Theorem 5.2. *Löb's theorem holds for the internal copy of* CA *and* PA *within* $\mathcal{C}[\mathsf{CA}]$ *or* $\mathcal{C}[\mathsf{PA}]$.

Proof. Again by similar arguments as in [13]. □

Finally, to come back to the logical practice, we should verify that Gödel's "standard way" of arithmetisation of PA should indeed result in the internal PA in our sense, i.e. the initial Boolean category with PNNO. We leave this for future work.

References

[1] Aczel, P. (1986). The type theoretic interpretation of constructive set theory: inductive definitions. In *Studies in Logic and the Foundations of Mathematics*, volume 114, pages 17–49. Elsevier.

[2] Feferman, S. (1960). Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92.

[3] Gambino, N. and Aczel, P. (2006). The generalised type-theoretic interpretation of constructive set theory. *The Journal of Symbolic Logic*, 71(1):67–103.

[4] Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik*, 38(1):173–198. Reprinted and translated in [5], pp. 144–195.

[5] Gödel, K. (1986). *Collected Works. Vol. 1: Publications 1929–1936*. Oxford University Press, Oxford.

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

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

[8] Hofmann, M. and Hofmann, M. (1997). Syntax and semantics of dependent types. *Extensional Constructs in Intensional Type Theory*, pages 13–54.

[9] Jacobs, B. (1999). *Categorical logic and type theory*. Elsevier.

[10] Johnstone, P. T. (2002). *Sketches of an Elephant: A Topos Theory Compendium*, volume 1, 2. Oxford University Press.

[11] Joyal, A. and Tierney, M. (1984). *An extension of the Galois theory of Grothendieck*, volume 309. American Mathematical Soc.

[12] Maietti, M. E. (2003). Joyal's arithmetic universes via type theory. *Electronic Notes in Theoretical Computer Science*, 69:272–286.

[13] van Dijk, J. and Oldenziel, A. G. (2020). Gödel incompleteness through arithmetic universes after a. joyal. *arXiv preprint arXiv:2004.10482*.

[14] Ye, L. (2023). Categorical structure in theory of arithmetic. *arXiv preprint arXiv:2304.05477*.