**RESEARCH ARTICLE**

# Finitely axiomatized theories lack self-comprehension

## Fedor Pakhomov[1,2]  |  Albert Visser[3]

[1]Vakgroep Wiskunde: Analysis, Logic and Discrete Mathematics, Ghent University, Ghent, Belgium

[2]Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia

[3]Philosophy, Faculty of Humanities, Utrecht University, Utrecht, The Netherlands

**Correspondence**
Fedor Pakhomov, Vakgroep Wiskunde: Analysis, Logic and Discrete Mathematics, Ghent University, Krijgslaan 281, Ghent B9000, Belgium.
Email: fedor.pakhomov@ugent.be

## Abstract

In this paper, we prove that no consistent finitely axiomatized theory one-dimensionally interprets its own extension with predicative comprehension. This constitutes a result with the flavor of the Second Incompleteness Theorem whose formulation is completely arithmetic-free. Probably the most important novel feature that distinguishes our result from the previous results of this kind is that it is applicable to arbitrary weak theories, rather than to extensions of some base theory. The methods used in the proof of the main result yield a new perspective on the notion of sequential theory, in the setting of forcing-interpretations.

**MSC 2020**
03F25 (primary), 03F40, 03E40 (secondary)

## 1 | INTRODUCTION

Gödel's incompleteness theorems [8] are probably the most widely known results in the field of Mathematical Logic. One reading of the First Incompleteness Theorem is just that, for a suitable theory, it produces an arithmetically true non-provable sentence or, under slightly stronger conditions, an independent sentence. This reading makes the First Incompleteness Theorem an extensional result.[†] The Second Incompleteness Theorem is more intricate. It refers to the formalized consistency statement and it is, thus, an intensional result.

Gödel's great 1931 paper was called *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.*[‡] The 'und verwandter Systeme' hints at the generality of what

---

[†] Of course, one also sees formulations of the theorem that refer explicitly to the specific sentences produced by Gödel. Under this guise, the theorem is *intensional*.

[‡] The English title is *On formally undecidable propositions of Principia Mathematica and related systems I*

> ### What is an interpretation?
>
> An $n$-dimensional interpretation of a first-order model $\mathfrak{A}$ in a first-order model $\mathfrak{B}$ is an isomorphic copy $\mathfrak{A}'$ of $\mathfrak{A}$ such that the domain $A'$ of $\mathfrak{A}'$ is a $\mathfrak{B}$-definable set of $n$-tuples of elements of $\mathfrak{B}$ and the evaluations of all $\mathfrak{A}'$ predicates and functions are $\mathfrak{B}$-definable. For example, we can two-dimensionally interpret the Euclidian plane in the field of reals by interpreting points on the plain by pairs of coordinates.
>
> An interpretation $K$ of a theory $T$ in a theory $U$ is a uniform construction of interpretation of models $K(\mathfrak{A})$ of $T$ inside models $\mathfrak{A}$ of $U$; here uniformity means that the first-order definitions of the domain and evaluations of non-logical symbols are the same for all models $\mathfrak{A}$. In particular, the example above naturally generalizes to an interpretation of axiomatic Euclidean geometry in the theory of real closed fields (ordered fields, where there are square roots of all positive numbers and all polynomials of odd degrees have roots).
>
> An important feature of interpretations is that they naturally give rise to straightforwardly defined syntactical translations $\tau$ from the language of $T$ to the language of $U$. The key property of the translation is that for a sentence $\varphi$ of the language of $T$ we have $\mathfrak{A} \vDash \varphi^\tau$ if and only if $K(\mathfrak{A}) \vDash \varphi$, for any model $\mathfrak{A}$ of $U$.
>
> We refer the reader to [16] or [17] for more details.

Gödel is doing. His argument will work for any system like Principia. In hindsight, the title still undersells what is accomplished. Gödel produced a uniform method to transform a suitable presentation of the axiom set of a theory into a sentence that, under certain specified circumstances, is independent of the theory. Later research established that we can make the conditions under which Gödel's argument works quite weak. Thus, the class of theories to which the argument is applicable is quite large. One form of the Second Incompleteness Theorem is the *no-interpretation version*.[†] We briefly explain interpretations in the minipage 'What is an interpretaation' and the the idea of the no-interpretation version in 'The No-Interpretation Version of the Second Incompleteness Theorem'. The results in this paper share the spirit of the *no-interpretation version* of the second theorem.

There are many good treatments of the incompleteness theorems nowadays in logic textbooks. The classical text on the intensionality of the second theorem is [5]. Good surveys are [11] and [3]. For treatments of the theorems in the context of weak theories, see [4] and [9]. A good and entertaining book on all misrepresentations of the theorems and fallacies surrounding the theorems is [7].

In this paper, we provide an impossibility argument in the niche of the Second Incompleteness Theorem. We show that no consistent finitely axiomatized theory $T$ can one-dimensionally interpret its own extension, $\mathsf{PC}(T)$ that is the second-order extension of $T$ by the predicative comprehension principle

$$\exists X \, \forall y \, (y \in X \leftrightarrow \varphi(y)),$$

where $\varphi$ has no second-order quantifiers and $X \notin \mathsf{FV}(\varphi)$.

---

[†] This name was coined by Harvey Friedman.

> ### The No-Interpretation Version of the Second Incompleteness Theorem
>
> A popular formulation of the Second Incompleteness Theorem is: *no consistent theory that is rich enough proves its own consistency*. So, what is a consistency statement? One answer, provided by Sol Feferman, in his paper [5], is that we fix an arithmetization of consistency that we recognize as intensionally correct. This sidesteps the difficult problem to say what a consistency statement is in general. You know it when you see it. A problem with Feferman's strategy is that it makes the consistency statements depend on a specific language. The solution to this problem is to replace provability by interpretability. A second question is: *what is sufficiently strong?* The development of arithmetics for computational complexity provide a good answer here: the theory should contain the weak arithmetic $S_2^1$ developed by Samuel Buss. See the minipage 'Buss' Theory $S_2^1$'.
>
> After these preparations, we can understand the formulation of the no-interpretation version. It says: no theory interprets $S_2^1$ plus its own consistency. Here the consistency statement is treated in the Feferman-style. We note that the *sufficiently strong* dropped out of the formulation: this is just because we made $S_2^1$, in a sense, part of the consistency statement.

Our result is inspired by well-known results about the connection between Predicative Comprehension and consistency in the case of sequential theories, roughly, theories with sufficient coding machinery. Two salient results are that first-order Peano Arithmetic, PA, does not interpret $ACA_0$ and that Zermelo–Fraenkel Set Theory, ZF, does not interpret Gödel–Bernays Set Theory, GB. One way of proving these results is to employ Second Incompleteness Theorem for extensions of the theory $S_2^1$ (see the minipage 'Buss' Theory $S_2^1$'). There is a result that $ACA_0$ is mutually interpretable with $S_2^1 + Con(PA)$ and GB is mutually interpretable with $S_2^1 + Con(ZF)$. We refer the reader to the minipage '$ACA_0$ and GB' for a brief description of $ACA_0$ and GB and to [14] for the details of this kind of mutual-interpretability result. Thus if PA would interpret $ACA_0$, then $S_2^1 + Con(PA)$ would prove $Con(ACA_0)$ and hence $Con(S_2^1 + Con(PA))$, contradicting Second Incompleteness Theorem (the argument for ZF and GB is the same).

More generally, we have the following result. Suppose $U$ is a sequential theory that is axiomatized by a scheme $\Theta$. Let $PC^{schem}(\Theta)$ be the theory obtained by taking Predicative Comprehension over the signature of $U$ and adding the universally quantified version of $\Theta$, where the schematic variables are replaced by class variables. (See the minipage 'Predicative Comprehension' for more information.) We have: $PC^{schem}(\Theta)$ is mutually interpretable with $S_2^1 + Con(\Theta)$. (See [15], for more information.) Note that $ACA_0 = PC^{schem}(PA)$ and $GB = PC^{schem}(ZF)$ (more precisely equality holds for the variant of GB given in two-sorted language). In combination with an appropriate version of the Second Incompleteness Theorem, we find that (†) $U$ does not interpret $PC^{schem}(\Theta)$.

In our paper, we study (†) outside of its comfort zone of sequential theories. We restrict ourselves to finitely axiomatized theories and to one-dimensional interpretability. However, under these restrictions, we prove the result for all theories. We note that, in the finitely axiomatized case, we only need $PC(U)$, the result of simply adding Predicative Comprehension to $U$, in stead of the more fancy $PC^{schem}(U)$.

The result that $T$ does not one-dimensionally interpret $PC(T)$ shows that $T$ does not interpret $S_2^1 + Con(T)$, since $S_2^1 + Con(T)$ interprets $PC(T)$. The argument for the interpretability of $PC(T)$

### $ACA_0$ and GB

The theory $ACA_0$ of Arithmetical Comprehension is a system of second-order arithmetic, that is, it talks about natural numbers and classes of natural numbers, whose set-existence principle (arithmetical comprehension) allows us to construct arbitrary sets of naturals defined by formulas without class-quantifiers.[†] It is one of central systems investigated in the field of Reverse Mathematics that aims to measure the logical strength of classical mathematical theorems (see [10]). The set of first-order consequences of $ACA_0$ precisely coincide with the consequences of PA.

Gödel–Bernays Set Theory GB is a set theory that, in addition to ordinary sets, also considers classes (that is, collections that are not members of other collections). The set of consequences of GB about ordinary sets coincides with Zermelo Fraenkel Set Theory ZF. In particular, GB is useful in formalization of arguments that manipulate with class-size collections of object such as category-theoretic arguments using large categories.

### Buss' Theory $S_2^1$

The theory $S_2^1$ was introduced by Sam Buss in his thesis [4] as a tool for the study of computational complexity. The provably recursive functions of $S_2^1$ are precisely the polynomial time computable functions. The arithmetization of syntax in $S_2^1$ is matter of course. As a consequence, the verification of the incompleteness theorems can be executed in $S_2^1$ without essential modifications. The theory $S_2^1$ is finitely axiomatizable. Moreover, $S_2^1$ is a *weak theory* in the sense that is mutually interpretable with theories like Robinson's Q and Adjunctive Set Theory AS.

in $S_2^1 + \mathsf{Con}(T)$ is essentially a refinement of the proof of the Completeness Theorem and does not involve diagonalization. Thus, for a restricted class of cases, our result implies a version of the Second Incompleteness Theorem in a diagonalization-free way.

One kind of advantage of our result over the no-interpretation version of the Second Incompleteness Theorem is that its formulation is arithmetization-free. (We note, however, that our proof is not arithmetization-free.) Thus, no arbitrary coding choices are needed to make our result specific. A second advantage is that, even if the earlier insight is perfectly general, this seems trifling in case $T$ does not interpret $S_2^1$. For example, the theory of the ordering of the natural numbers, say $W$, is finitely axiomatizable and decidable. The theory $S_2^1$ is essentially undecidable. So, $W$ does not interpret $S_2^1$. So, it is not informative to know that $W$ does not interpret $S_2^1 + \mathsf{Con}(W)$. On the other hand, it is informative that $W$ does not one-dimensionally interpret $\mathsf{PC}(W)$.

We note that, for number of specific non-sequential finitely axiomatizable theories, one can prove that they do not one-dimensionally interpret their own predicative comprehension relatively easily by various ad hoc arguments. For example, for theories $T$ having finite models, this can be proved using the fact that the smallest finite model of $T$ will be smaller than the smallest

---

[†] This form of comprehension is also known as predicative comprehension. See the minipage 'Predicative Comprehension'.

> **Predicative Comprehension**
>
> In Predicative Comprehension, we limit the comprehension axiom to formulas that only contain quantifiers ranging over objects and not over classes. The idea is that defining a class using quantification over a totality that the class in question belongs to involves a circle. The demand for the exclusion of such vicious circularity was first advocated by Henri Poincaré.
>
> It turns out that the operation of adding predicative comprehension to a theory has many good properties. It is, so to speak, a natural thing to do independent of philosophical motivations. This operation (and closely connected operations) have functorial or functor-like properties as will be discussed in the present paper. In case the given theory has pairing, one can show that predicative comprehension is finitely axiomatizable over it. If the theory has good sequence coding, it turns out that predicative comprehension has an intimate connection to consistency statements. This observation is the starting point of the work in the present paper, where we think of predicative comprehension as a consistency analogue.

finite model of $PC(T)$. For Q, one can use a first-order definable $S_2^1$-cut to show that $PC(Q)$ one-dimensionally interprets $PC(S_2^1)$, thus reducing non-interpretability of predicative comprehension to the case of the sequential theory $S_2^1$. Although we have not checked it carefully, it appears that, for the theory of dense linear orders DLO, it is possible to use complexity theoretic reasoning: the theory itself is known to be PSPACE-complete [6] and it seems that, using fairly standard techniques, one can show that any consistent extension of $PC(DLO)$ is $DTime(2^{2^{x^{o(1)}}})$-hard. However, these diverse proofs fail to be uniform among different kinds of theories, whereas our result establishes a fully general phenomenon.

Our paper provides some spin-offs that hold independent interest. We present these results in Section 7.

A first spin-off result tells us that the extension of a theory $T$ with adjunctive sets is mutually forcing-interpretable with the extension of $T$ with the adjunctive theory of binary relation classes plus the no-universe axiom. The result has the extra feature that the forcing-interpretations back-and-forth preserve the objects and relations of $T$. We note that adding adjunctive sets is a form of sequential closure, that is, a way of making a theory sequential.

A second result tells us that, if $T$ is finitely axiomatized and one dimensionally interprets $T$ on a provably smaller domain, then the extension of $T$ with $n$-ary adjunctive classes, for sufficiently large $n$, forcing-interprets the extension of $T$ with adjunctive sets.

Third, we show that, if $T$ is finitely axiomatized and one dimensionally interprets the extension of $T$ with adjunctive classes, then $T$ forcing-interprets the extension of $T$ with adjunctive sets.

## Genesis of this work

The questions leading to the results of this paper come from earlier work by Albert Visser. The strengthening of the non-interpretability result of $PC(T)$ in $T$ for the sequential, finitely

axiomatized case, to the case of pairing theories was found in a conversation of Albert Visser and Fedor Pakhomov. The basic proof strategy for Theorem 1 was discovered by Fedor Pakhomov.

## 2 | PRELIMINARIES

All theories that we consider are one-sorted theories with equality and finite relational signature. We assume that the connectives in the first-order language are $\forall, \wedge$, and $\neg$. We express all the other connectives using these ones.

However, we frequently will consider theories that naturally should be considered $n$-sorted theories (with relational signature). In order to do this, we will identify an $n$-sorted theory $T$, whose sorts are $\sigma_1, \dots, \sigma_n$ with the following one-sorted theory $T^\flat$. The signature of $T^\flat$ contains, in addition to the signature of $T$, unary predicate symbols $\mathsf{S}_{\sigma_1}, \dots, \mathsf{S}_{\sigma_n}$. We consider the sorted quantifier $\forall x^{\sigma_i} \varphi$ to be a shorthand for $\forall x \, (\mathsf{S}_{\sigma_i}(x) \to \varphi)$. In addition to the explicitly given axioms of $T$, we have the following axioms.

(1) $\bigvee_{1 \leqslant i \leqslant n} \mathsf{S}_{\sigma_i}(x)$.

(2) $\neg (\mathsf{S}_{\sigma_i}(x) \wedge \mathsf{S}_{\sigma_j}(x))$, for $i < j$.

(3) $\exists x \, \mathsf{S}_{\sigma_i}(x)$, for each $i$.

(4) $R(x_1, \dots, x_m) \to (\mathsf{S}_{\sigma_{k_1}}(x_1) \wedge \dots \wedge \mathsf{S}_{\sigma_{k_m}}(x_m))$, for each original $m$-ary predicate symbol $R$, whose $i$th argument is of the sort $\sigma_i$, for $i < m$.

Here we treat identity separately: identity of each sort is simply the restriction of identity for the whole domain of $T^\flat$ to each of the domains $\mathsf{S}_\sigma$.

For theories $T$ and $U$, we denote as $T \sqcup U$ the two-sorted theory that has all predicates of $T$ on the first sort, all predicates of $U$ on the second sort, and whose axioms are all the axioms of $T$ relativized to the first sort and all the axioms of $U$ relativized to the second sort.

We define theory $\mathsf{PC}_{\leqslant n}(T)$ (Predicative Comprehension up to the arity $n$), for any theory $T$. This is the $n + 1$ sorted theory, whose sorts are $\mathfrak{o}$ and $\mathfrak{c}_1, \dots, \mathfrak{c}_n$. The predicates of $\mathsf{PC}_{\leqslant n}(T)$ are the predicates of $T$ restricted to the sort $\mathfrak{o}$ as well as the predicates $\langle x_1^\mathfrak{o}, \dots, x_i^\mathfrak{o} \rangle \in X^{(i)}$, for $1 \leqslant i \leqslant n$. The axioms of $\mathsf{PC}_{\leqslant n}(T)$ are as follows.

(1) The axioms of $T$ relativized to the sort $\mathfrak{o}$.

(2) $\exists X^{\mathfrak{c}_k} \forall x_1^\mathfrak{o}, \dots, x_k^\mathfrak{o} \, (\langle x_1, \dots, x_k \rangle \in X \leftrightarrow \varphi(x_1, \dots, x_k))$, where all quantifiers in $\varphi$ are on the sort $\mathfrak{o}$ and $\varphi$ could contain additional free variables other than $X$ of all sorts.

(3) $\forall X^{\mathfrak{c}_k}, Y^{\mathfrak{c}_k} \, (\forall x_1^\mathfrak{o}, \dots, x_k^\mathfrak{o} \, (\langle x_1, \dots, x_k \rangle \in X \leftrightarrow \langle x_1, \dots, x_k \rangle \in Y) \to X = Y)$.

The theory $\mathsf{PC}(T)$ is $\mathsf{PC}_{\leqslant 1}(T)$.

In this paper, we consider multi-dimensional relative interpretations with parameters and definable equality. When we talk about $n$-dimensional interpretations, we mean interpretations whose domain is $n$-dimensional and that could use parameters and definable equality. See, for example, [16] or [17] for definitions.

Our main theorem is

**Theorem 1.** *No consistent finitely axiomatized theory $T$ can one-dimensionally interpret* $\mathsf{PC}(T)$ *In other words, for every consistent finitely axiomatized theory $T$ we have* $T \not\vdash_1 \mathsf{PC}(T)$.

## 3 │ PREDICATIVE COMPREHENSION AND TUPLES

We write $T \rhd_m U$ if $T$ interprets $U$ by an $m$-dimensional interpretation. We have the following trivial lemma.

**Lemma 1.** *If $T \rhd_1 U$, then $\mathsf{PC}(T) \rhd_1 \mathsf{PC}(U)$.*

And its multi-dimensional generalization:

**Lemma 2.** *If $T \rhd_n U$, then $\mathsf{PC}_{\leqslant nm}(T) \rhd_n \mathsf{PC}_{\leqslant m}(U)$.*

It is sometimes pleasant to treat dimension using an auxiliary theory that adds $i$-tuples for $2 \leqslant i \leqslant n$ to the given base theory. Let $\mathsf{Tuple}_{\leqslant n}(T)$ be the following $n$-sorted theory. The sorts of $\mathsf{Tuple}_{\leqslant n}(T)$ are $\mathfrak{t}_1, \dots, \mathfrak{t}_n$. Here $\mathfrak{t}_1$ may be identified with $\mathfrak{o}$, the sort of basic objects. The signature of $\mathsf{Tuple}_{\leqslant n}$ consists of all the predicates of $T$ on the sort $\mathfrak{t}_1$ and the predicates $\mathsf{Tp}_i(p^{\mathfrak{t}_i}, x_1^{\mathfrak{t}_1}, \dots, x_i^{\mathfrak{t}_1})$, for all $2 \leqslant i \leqslant n$. The axioms of $\mathsf{Tuple}_{\leqslant n}(T)$ are:

(1) all the axioms of $T$ relativized to $\mathfrak{t}_1$;
(2) $\forall p^{\mathfrak{t}_i}, q^{\mathfrak{t}_i}, x_1^{\mathfrak{t}_1}, \dots, x_i^{\mathfrak{t}_1}, y_1^{\mathfrak{t}_1}, \dots, y_i^{\mathfrak{t}_1}((\mathsf{Tp}_i(p, x_1, \dots, x_i) \wedge \mathsf{Tp}_i(q, y_1, \dots, y_i)) \rightarrow (p = q \leftrightarrow (x_1 = y_1 \wedge \dots \wedge x_i = y_i)))$, for $2 \leqslant i \leqslant n$;
(3) $\forall p^{\mathfrak{t}_i} \exists x_1^{\mathfrak{t}_1}, \dots, x_i^{\mathfrak{t}_1} \, \mathsf{Tp}_i(p, x_1, \dots, x_i)$, for $2 \leqslant i \leqslant n$;
(4) $\forall x_1^{\mathfrak{t}_1}, \dots, x_i^{\mathfrak{t}_1} \exists p^{\mathfrak{t}_i} \, \mathsf{Tp}_i(p, x_1, \dots, x_i)$, for $2 \leqslant i \leqslant n$.

We denote as $\mathsf{PC}^n(T)$ the $n$ time application of the PC-operator to the theory $T$, that is, $\mathsf{PC}^0(T) = T$ and $\mathsf{PC}^{n+1}(T) = \mathsf{PC}(\mathsf{PC}^{n+1}(T))$.

**Lemma 3.** $\mathsf{PC}^2(T) \rhd_1 \mathsf{Tuple}_{\leqslant 2}(T)$.

*Proof.* Theory $\mathsf{PC}^2(T)$ is a theory that may be considered to be 3-sorted: we have the sort of elements (on which we have $T$), the sort of classes of elements, and the sort of classes that could contain either elements or other classes of elements. We represent pairs $\langle a, b \rangle$ by Kuratowski-style pairs $\{\{a\}, \{a, b\}\}$ (in the domain of classes of classes) and we represent elements by themselves. The verification of all axioms of $\mathsf{Tuple}_{\leqslant 2}(T)$ is routine. □

Trivially we have:

**Lemma 4.** $\mathsf{PC}(\mathsf{Tuple}_{\leqslant n}(T)) \rhd_1 \mathsf{PC}_{\leqslant n}(T)$.

**Lemma 5.** $\mathsf{Tuple}_{\leqslant n}(\mathsf{Tuple}_{\leqslant m}(T)) \rhd_1 \mathsf{Tuple}_{\leqslant nm}(T)$.

*Proof.* In $\mathsf{Tuple}_{\leqslant n}(\mathsf{Tuple}_{\leqslant m}(T))$, we have $T$-domain, tuples of the elements of $T$-domain $\langle a_1, \dots, a_k \rangle_1$, where we have $1 \leqslant k \leqslant m$ and the tuples $\langle s_1, \dots, s_r \rangle_2$, where $1 \leqslant r \leqslant n$ and each component $s_i$ is either element of $T$-domain or tuples $\langle a_1, \dots, a_k \rangle_1$. Our interpretation preserves $T$-domain and all $T$ predicates. We represent a tuple $\langle a_1, \dots, a_p \rangle$, for $1 \leqslant p \leqslant nm$ as follows. We find the unique $0 \leqslant r < m$ and $1 \leqslant \ell \leqslant n$ such that $p = rn + \ell$ and put our representation to be $\langle s_1, \dots, s_{r+1} \rangle_2$, where, for $1 \leqslant i \leqslant r$, we put $s_i = \langle a_{(i-1)r+1}, \dots, a_{(i-1)r+n} \rangle_2$ and we take $s_{r+1} = \langle a_{rn+1}, \dots, a_{rn+\ell} \rangle_1$. □

From Lemmas 1, 3, and 5, we get

**Lemma 6.** $\mathsf{PC}^{2n}(T) \rhd_1 \mathsf{Tuple}_{\leqslant 2^n}(T)$.

Combining Lemmas 1, 4, and 6, we get

**Lemma 7.** $\mathsf{PC}^{2n+1}(T) \rhd_1 \mathsf{PC}_{\leqslant 2^n}(T)$.

## 4 │ FORCING SEQUENTIALITY

In addition to the usual kinds of interpretations, we consider *forcing-interpretations* (see the survey by Avigad [2] for an overview of the method).

For a theory $T$, let $\mathsf{KM}(T)$ (Kripke models of $T$) be the following two-sorted theory. The sorts of $\mathsf{KM}(T)$ are:

(1) $\mathfrak{w}$ (sort of worlds);
(2) $\mathfrak{d}$ (sort of elements of domains in worlds).

The relations $\mathsf{KM}(T)$ are:

(1) the binary predicate $p^{\mathfrak{w}} \preceq q^{\mathfrak{w}}$ (accessibility relation on worlds);
(2) binary predicate $D(p^{\mathfrak{w}}, x^{\mathfrak{d}})$ (for a fixed $p$ it defines the domain $D_p$ of the Kripke model in the world $p$);
(3) the predicate $R^{\star}(p^{\mathfrak{w}}, x_1^{\mathfrak{d}}, \dots, x_k^{\mathfrak{d}})$ for each $k$-ary predicate $R$ of the signature of $T$ (for each fixed $p$ it gives the interpretation of $R$ in the world $p$).

For each formula $\varphi(x_1, \dots, x_n)$ of the language of $T$, we define by recursion the formulas $p \Vdash \varphi(x_1, \dots, x_n)$ (the model forces $\varphi$ in the world $p$) of the language of $\mathsf{KM}(T)$.

(1) $p \Vdash R(x_1, \dots, x_n)$ is $(\forall q^{\mathfrak{w}} \preceq p)(\exists r^{\mathfrak{w}} \preceq q)\, R^{\star}(r, x_1, \dots, x_n)$.
(2) $p \Vdash \varphi(x_1, \dots, x_n) \wedge \psi(x_1, \dots, x_n)$ is $(p \Vdash \varphi(x_1, \dots, x_n)) \wedge (p \Vdash \psi(x_1, \dots, x_n))$.
(3) $p \Vdash \neg\, \varphi(x_1, \dots, x_n)$ is $(\forall q^{\mathfrak{w}} \preceq p) \neg\, (q \Vdash \varphi(x_1, \dots, x_n))$.
(4) $p \Vdash \forall y\, \varphi(x_1, \dots, x_n, y)$ is $(\forall q^{\mathfrak{w}} \preceq p)(\forall y^{\mathfrak{d}})(D(q, y) \to q \Vdash \varphi(x_1, \dots, x_n, y))$.

The axioms of $\mathsf{KM}(T)$ are:

(1) $\forall p^{\mathfrak{w}}\, p \preceq p$ (reflexivity of $\preceq$).
(2) $\forall q^{\mathfrak{w}}, q^{\mathfrak{w}}, r^{\mathfrak{w}}((p \preceq q \wedge q \preceq r) \to p \preceq r)$ (transitivity of $\preceq$).
(3) $\forall p^{\mathfrak{w}} \exists x^{\mathfrak{d}}\, D(p, x)$ (domains are not empty).
(4) $\forall p^{\mathfrak{w}}, q^{\mathfrak{w}}\, (q \preceq p \to \forall x^{\mathfrak{d}}(D(p, x) \to D(q, x)))$ ($D_p \subseteq D_q$, for $q \preceq p$).
(5) $\forall p^{\mathfrak{w}}, x_1^{\mathfrak{d}}, \dots, x_k^{\mathfrak{d}}\, (R^{\star}(p, x_1, \dots, x_k) \to (D(p, x_1) \wedge \dots \wedge D(p, x_k)))$.
(6) $\forall p^{\mathfrak{w}}, q^{\mathfrak{w}}\, (q \preceq p \to \forall x_1^{\mathfrak{d}}, \dots, x_k^{\mathfrak{d}}\, (R^{\star}(p, x_1, \dots, x_k) \to R^{\star}(q, x_1, \dots, x_k))$ (downward persistence of the interpretations of predicates).
(7) $\forall p^{\mathfrak{w}}\, p \Vdash \varphi$, for all axioms $\varphi$ of $T$.

We say that $U$ is forcing-interpretable in $T$ if there is an interpretation of $\mathsf{KM}(U)$ in $T$.

Note that, usually, Kripke models are formulated for intuitionistic predicate logic rather than classical predicate logic. However, in fact, the Kripke models as defined above always force all tautologies of classical predicate logic and not just tautologies of intuitionistic logic. We made

two design choices to ensure this property. First, we restrict our connectives to just $\wedge, \neg, \forall$. And second our definition of forcing of atomic formulas differs from the usual one: for forceability of $R(x_1, \ldots, x_n)$ in a world $p$, we demand that the points where $R(x_1, \ldots, x_n)$ is valid are dense below $p$, whereas the usual definition would simply demand that $R(x_1, \ldots, x_n)$ should be valid in $p$. That is, our definition of forceability of atomic formulas $R(x_1, \ldots, x_n)$ is equivalent to forceability of $\neg\neg R(x_1, \ldots, x_n)$ in the usual sense. It is well known that the fragment of intuitionistic predicate logic that is restricted to the connectives $\wedge, \neg, \rightarrow, \forall$ and where double negations are put over all atomic formulas (negative fragment) in fact coincides with the restriction of classical predicate logic to the same class of formulas. See, for example, [12, Chapter 2, Section 3].

Immediately from the definition of forcing-interpretation and the fact that interpretations are closed under composition, we get

**Lemma 8.** *If $T$ interprets $U$ and $U$ forcing-interprets $V$, then $T$ forcing-interprets $V$.*

*Remark* 1. Although we have not checked this carefully, it appears that it is possible to compose forcing-interpretations (and hence forcing-interpretability is a pre-order). However, we do not need this fact to obtain the results of the present paper. We note that it is likely that composition of forcing-interpretations will raise the dimension of the composition.

**Lemma 9.** *There is an interpretation of $\mathsf{KM}(\mathsf{PC}(T))$ in $\mathsf{PC}_{\leqslant 2}(\mathsf{KM}(T))$.*

*Proof.* We work in $\mathsf{PC}_{\leqslant 2}(\mathsf{KM}(T))$ to define the desired interpretation.

We already have an internal Kripke model $\mathcal{K}$ of $T$ inside the $\mathfrak{o}$-sort. That is, we have a poset of worlds $P^{\mathcal{K}}$, a family of domains $\langle D_p^{\mathcal{K}} \mid p \in P \rangle$ and interpretations $\langle R_p^{\mathcal{K}} \mid p \in P \rangle$ of all $T$ predicates $R$.

We define a Kripke model $S$ of $\mathsf{PC}(T)$. The poset of worlds $P^S$ simply coincides with $P^{\mathcal{K}}$. We call a $\mathfrak{c}_2$-set $A$ a *name* if it consists only of pairs $\langle p, x \rangle$ such that $p \in P^{\mathcal{K}}$ and $x \in D_p$. For each world $p$, the domain $D_p^S$ extends the domain $D_p^{\mathcal{K}}$ by all names.

Consider a world $p$.

(1) We put $S, p \Vdash S_{\mathfrak{o}}(x)$ if and only if $x \in D_p^S$.
(2) We put $S, p \Vdash S_{\mathfrak{c}_1}(A)$ if and only if $A$ is a name.
(3) For each $k$-ary predicate $R$ of $T$ and $x_1, \ldots, x_k \in D_p^S$ we put $S, p \Vdash R(x_1, \ldots, x_k)$ if and only if $x_1, \ldots, x_k \in D_p^{\mathcal{K}}$ and $\mathcal{K}, p \Vdash R(x_1, \ldots, x_k)$.
(4) We put $S, p \Vdash x \in A$ if and only if $x \in D_p^{\mathcal{K}}$, $A$ is a name and there exists $q \geq p$ such that $\langle q, x \rangle \in A$.

We note that the downward persistence of $\in$ is guaranteed by the definition. The fact that $\mathcal{K}$ forces the axioms of $T$ obviously implies that $S$ forces the relativizations to $\mathfrak{o}$ of the axioms of $T$. Let us verify in a world $p$ the forceability of an instance of predicative comprehension

$$\exists X^{\mathfrak{c}_1} \forall x^{\mathfrak{o}}(x \in X \leftrightarrow \varphi(x, \vec{a}, \vec{A})), \text{ where } \vec{a} \in D_p^{\mathcal{K}} \text{ and } \vec{A} \text{ are names.}$$

Let $B$ be the following name:

$$B = \{\langle q, y \rangle \mid q \leq p, y \in D_q^{\mathcal{K}}, \text{ and } S, q \Vdash \varphi(y, \vec{a}, \vec{A})\}.$$

The definition is correct (that is, we obtain $B$ by predicative comprehension), since $\varphi$ does not have quantifiers over classes and, thus, $S, q \Vdash \varphi(y, \vec{a}, \vec{A})$ is also expressible by a formula without quantifiers over classes. It is easy to see that the formula $\forall x^{\mathfrak{o}}(x \in B \leftrightarrow \varphi(x, \vec{a}, \vec{A}))$ is forced in $p$.

We did not yet treat identity of classes, but that can be easily added by setting $p \Vdash A = B$ if and only if $p \Vdash \forall z \, (z \in A \leftrightarrow z \in B)$. □

**Corollary 1.** *If there is a forcing-interpretation of $U$ in $T$, then there is a forcing-interpretation of* $\mathsf{PC}(U)$ *in* $\mathsf{PC}_{\leqslant n}(T)$*, for some $n$.*

*Proof.* In view of Lemma 9, it is sufficient to define an interpretation of the theory $\mathsf{PC}_{\leqslant 2}(\mathsf{KM}(U))$ in $\mathsf{PC}_{\leqslant n}(T)$. The latter can be done using Lemma 2. □

For a theory $T$, we denote as $\mathsf{AS}(T)$ (Adjunctive set theory) the extension of $T$ by a fresh predicate symbol $x \in y$ and axiom.

(1) $\exists x \forall y \, \neg \, y \in x$.
(2) $\exists z \forall w \, (w \in z \leftrightarrow (w \in x \vee w = y))$.

A theory $T$ is called *sequential* if it admits a definitional extension to $\mathsf{AS}(T)$.

**Lemma 10.** *Suppose $T$ is finitely axiomatized theory such that there is a one-dimensional interpretation of $T \sqcup \forall x \, (x = x)$ in $T$. Then there is a forcing-interpretation of $\mathsf{AS}(T)$ in $\mathsf{PC}_{\leqslant n}(T)$, for sufficiently large $n$.*

*Proof.* Let $n$ be the maximum of the arities of all predicates in $T$. We have $n \geqslant 2$, since we have equality in the signature of $T$. We work in $\mathsf{PC}_{\leqslant n}(T)$.

A model $M$ of the signature of $T$ is a tuple consisting of a $\mathfrak{c}_1$-class $D^M$ giving the domain of the model and $(k_R)$-classes $R^M$, for each $T$-predicate $R$ of arity $k_R$. Naturally, we express satisfaction of formulas inside $M$. We call $M$ a model of $T$ if all axioms of $T$ are satisfied in it. Note that here we do not require the absoluteness of equality, that is, the equality predicate $=^M$ is simply an equivalence relation. Note also that there is the model of $T$, whose domain is the whole $\mathfrak{o}$-sort and whose predicates are interpreted identically in $\mathsf{PC}_{\leqslant n}(T)$. Using the one-dimensional interpretation of $T \sqcup \forall x (x = x)$ in $T$, for any model $M$ of $T$ we obtain a model $M'$ of $T$ such that $D^M \supsetneq D^{M'}$.

We say that a $\mathfrak{c}_1$-class $A$ is *small* if there are no models $M \vDash T$ such that $D^M \subseteq A$. The class of the elements of the whole $\mathfrak{o}$-sort is not small, since there is the model of $T$, whose domain is the whole $\mathfrak{o}$-sort. Observe that, for any small $A$ with $x \notin A$, the class $A \cup \{x\}$ is also small. Otherwise, there would be a model $M$ of $T$, whose domain is contained in $A \cup \{x\}$, hence there would be a model $M'$ of $T$ with $D^{M'} \subsetneq A \cup \{x\}$ and, thus, either $M'$ itself, or the result of swapping some element in its domain with $x$, would be a model $M''$ of $T$, whose domain is contained in $A$, contradicting the smallness of $A$.

A binary relation $H$ is a pair consisting of a $\mathfrak{c}_1$-class $D^H$ and a $\mathfrak{c}_2$-class $R^H$ such that, whenever $\langle x, y \rangle \in R^H$, we have $x, y \in D^H$. We use $x R^H y$ as a shorthand for $\langle x, y \rangle \in R^H$. We say that a binary relation $H$ end-extends a binary relation $K$ and write $H \supseteq_{\mathrm{end}} K$, if

(1) $D^H \supseteq D^K$;
(2) for any $x, y \in D^K$ we have $x R^K y$ if and only if $x R^H y$;
(3) for any $x \in D^K$ and $y \in D^H \setminus D^K$ we have $\neg \, y R^H x$.

We say that a binary relation is *small* if its domain is a small $\mathfrak{c}_1$-class.

To finish the proof, we define an interpretation of $\mathsf{KM}(\mathsf{AS}(T))$. The poset of the worlds of the Kripke model consists of the small binary relations ordered by $\supseteq_{\mathsf{end}}$ (a small binary relation accesses all its small end-extensions). The domain in each world is simply the whole $\mathfrak{o}$-sort. The interpretations of all the predicates of $T$ in all the worlds are simply the classes corresponding to the predicates of $T$. Finally, we interpret the predicate $\in$ in the world $H$ as $R^H$.

It is trivial to see that, in the Kripke model thus defined, all the axioms of $T$ are forced. The forceability of the axiom of empty class $\exists x \, \forall y \, \neg \, y \in x$ is clearly equivalent to the following true statement: (†) for any world $H$, there is a world $K \supseteq_{\mathsf{end}} H$ and a $\mathfrak{o}$-object $x$, such that, for any $\mathfrak{o}$-object $y$ and $L \supseteq_{\mathsf{end}} K$, we have $\neg \, yR^L x$. The statement (†) is true since, for a given small binary relation $H$, we can take as $x$ any element outside of $D^K$, and define the small $K \supseteq_{\mathsf{end}} H$ with the domain $D^K = D^H \cup \{x\}$ so that $\neg \, yR^K x$, for any $y \in D^K$. We verify the axiom of adjunction $\exists z \, \forall w \, (w \in z \leftrightarrow (w \in x \lor w = y))$ in a similar manner. Thus, we indeed have defined an interpretation of $\mathsf{KM}(\mathsf{AS}(T))$. $\qquad \square$

*Remark* 2. We note that the forcing-interpretation defined in the proof of Lemma 10 is an analogue of what is called an $\mathfrak{o}$-direct interpretation in [13]. This means that the interpretation preserves the domain and the identity relation for the $\mathfrak{o}$-sort. Moreover, it preserves $T$ identically on the $\mathfrak{o}$-sort.

## 5 | PROOF OF THE MAIN THEOREM

Recall that $\mathsf{S}_2^1$ is a weak arithmetical system capable of the natural formalization of arguments about $P$-time computable functions (see, for example, [4]). We will assume that finitely axiomatized theories are given inside $\mathsf{S}_2^1$ with the obvious representations of their axiom set.

**Theorem 2** [14]. *For any finitely axiomatized sequential $T$, the theory $\mathsf{PC}(T)$ interprets $\mathsf{S}_2^1 + \mathsf{Con}(T)$.*

Since for finitely axiomatizable theories both interpretations and forcing-interpretations lead to natural $P$-time transformations of proofs in the interpreted theory to proofs in the interpreting theory we have the following lemma.

**Lemma 11.** *Suppose $T$ and $U$ are finitely axiomatized theories. If $T$ interprets $U$, then $\mathsf{S}_2^1 \vdash \mathsf{Con}(T) \to \mathsf{Con}(U)$. If $T$ forcing-interprets $U$, then we have $\mathsf{S}_2^1 \vdash \mathsf{Con}(T) \to \mathsf{Con}(U)$.*

*Proof.* The case of usual interpretations is well known so we will treat only the case of forcing-interpretations.

The forcing-interpretations correspond to polynomial-time transformations of proofs (see a discussion in [1, 2]). This enables us to formalize in $\mathsf{S}_2^1$ the following reasoning (since $\mathsf{S}_2^1$ is able to naturally work with the polynomial transformations of strings). To prove $\mathsf{Con}(T) \to \mathsf{Con}(U)$, we assume there is a proof $P$ of contradiction from axioms of $U$ and show that then there is a proof of contradiction from axioms of $T$. Indeed, using forcing-interpretation of $U$ in $T$, we simple transform $P$ to a $T$ proof of forceability of falsity, which leads to a proof of contradiction from the axioms of $T$. $\qquad \square$

**Theorem 3** (Gödel's Second Incompleteness for interpretations $\mathsf{S}_2^1$). *No consistent r.e. $T$ interprets $\mathsf{S}_2^1 + \mathsf{Con}(T)$.*

Finally, we remind the reader of a basic fact about PC, for a proof see, for example, [14].

**Lemma 12.** *Suppose $T$ is finitely axiomatized and sequential. Then,* $\mathsf{PC}(T)$ *is finitely axiomatizable.*

Now let us prove Theorem 1.

*Proof.* Assume for a contradiction that $T$ one-dimensionally interprets $\mathsf{PC}(T)$. We reason as follows using previously proven lemmas.

(1) $T$ one-dimensionally interprets $\mathsf{PC}^n(T)$, for any $n$ (by Lemma 1).

(2) $T$ one-dimensionally interprets $\mathsf{PC}_{\leqslant n}(T)$, for any $n$ (by 1 and Lemma 7).

(3) $T$ one-dimensionally interprets $T \sqcup \forall x(x = x)$ (this trivially follows from the fact that $T$ one-dimensionally interprets $\mathsf{PC}(T)$).

(4) $\mathsf{PC}_{\leqslant n}(T)$ forcing-interprets $\mathsf{AS}(T)$, for some $n$ (by 3 and Lemma 10).

(5) $\mathsf{PC}_{\leqslant m}(\mathsf{PC}_{\leqslant n}(T))$ forcing-interprets $\mathsf{PC}(\mathsf{AS}(T))$, for some $n$ and $m$ (by 4 and Corollary 1).

(6) $\mathsf{PC}^m(\mathsf{PC}_{\leqslant n}(T))$ forcing-interprets $\mathsf{PC}(\mathsf{AS}(T))$, for some $n$ and $m$ (by 5 and Lemma 7).

(7) $\mathsf{PC}^n(T)$ forcing-interprets $\mathsf{PC}(\mathsf{AS}(T))$, for some $n$ (by 6 and Lemma 8).

(8) $T$ forcing-interprets $\mathsf{PC}(\mathsf{AS}(T))$ (by 1, 7 and Lemma 8).

(9) $\mathsf{PC}(\mathsf{AS}(T))$ interprets $\mathsf{S}^1_2 + \mathsf{Con}(T)$ (by Theorem 2).

(10) $\mathsf{PC}(\mathsf{AS}(T))$ interprets $\mathsf{S}^1_2 + \mathsf{Con}(\mathsf{PC}(\mathsf{AS}(T)))$ (by 8 in combination with Lemmas 12 and 11).

(11) $\mathsf{PC}(\mathsf{AS}(T))$ is inconsistent (by 10, Theorem 3).

(12) $T$ is inconsistent (by 11 and 8).

So, we are done. □

## 6 | THE MULTI-DIMENSIONAL CASE

In this section, we sketch a proof of a generalization of Theorem 1

**Theorem 4.** *No consistent finitely axiomatized $T$ can $n$-dimensionally interpret* $\mathsf{PC}_{\leqslant n}(T)$.

Let us define the theory $T^n$ that is similar to the theory $\mathsf{Tuple}_{\leqslant n}$ but instead of having sorts of elements and sorts of $k$-tuples, $2 \leqslant k \leqslant n$, the theory $T^n$ is one-sorted and all elements are treated as $n$-tuples. The key property of $T^n$ is Lemma 13 that allows us to reduce the study of $n$-dimensional interpretations to the study of 1-dimensional interpretations (we note that the right to left direction of Lemma 13 would fail if we would replace $T^n$ with $\mathsf{Tuple}_{\leqslant n}$, since in general $\mathsf{Tuple}_{\leqslant n}$ is not $n$-dimensionally interpretable in $T$).

The signature of $T^n$ expands the signature of $T$ by a unary predicate $\mathsf{Dg}$ and and $n + 1$-ary predicate $\mathsf{Tp}$. The axioms of $T^n$ are:

(1) relativization of the axioms of $T$ to $\mathsf{Dg}$;

(2) $\forall x, y_1, \dots, y_n \, (\mathsf{Tp}(x, y_1, \dots, y_n) \to \bigwedge_{1 \leqslant i \leqslant n} \mathsf{Dg}(y_i))$;

(3) $\forall x, x', y_1, \dots, y_n, z_1, \dots, z_n$
$((\mathsf{Tp}(x, y_1, \dots, y_n) \wedge \mathsf{Tp}(x, z_1, \dots, z_n)) \to (x = x' \leftrightarrow \bigwedge_{1 \leqslant i \leqslant n} y_i = z_i))$;

(4) $\forall x \, \exists y_1, \dots, y_n \, \mathsf{Tp}(x, y_1, \dots, y_n)$;

(5) $\forall y_1, \dots, y_n \, ( \bigwedge_{1 \leqslant i \leqslant n} \mathsf{Dg}(y_i) \to \exists x \, \mathsf{Tp}(x, y_1, \dots, y_n));$

(6) $\forall x \, (\mathsf{Dg}(x) \to \mathsf{Tp}(x, \dots, x)).$

In $T^n$, we treat $x$ such that $\mathsf{Dg}(x)$ as individuals and we treat arbitrary objects $x$ as tuples of individuals ($x$ corresponds to the unique tuple $\langle y_1, \dots, y_n \rangle$ such that $\mathsf{Tp}(x, y_1, \dots, y_n)$).

It is easy to see that the following lemma holds:

**Lemma 13.** *There is an $n$-dimensional interpretation of $U$ in $T$ if and only if there is an one-dimensional interpretation of $U$ in $T^n$*

For a theory $T$, let us define the theory $\mathsf{PC}^{\mathsf{st}}(T)$. The language of $\mathsf{PC}^{\mathsf{st}}(T)$ extends the language of $T$ by a fresh unary predicate $\mathsf{Sng}$ and a binary predicate $\in$. The theory $\mathsf{PC}^{\mathsf{st}}(T)$ has the following axioms.

(1) The axioms of $T$ relativized to $\mathsf{Sng}$.
(2) $\forall x (\forall y (y \in x \leftrightarrow y = x) \leftrightarrow \mathsf{Sng}(x)).$
(3) $\forall \vec{p} \, \exists x \, \forall y \, (y \in x \leftrightarrow (\mathsf{Sng}(y) \wedge \varphi(y, \vec{p})))$, where $\varphi$ is a formula where all occurrences of quantifiers are of the form $\forall z \, (\mathsf{Sng}(z) \to \psi)$.

Note that the theory $\mathsf{PC}^{\mathsf{st}}(T^n)$ in effect is very similar to $\mathsf{PC}_{\leqslant n}(T)$. Namely, we can simulate, in $\mathsf{PC}^{\mathsf{st}}(T^n)$, the sort $\mathfrak{o}$ by $x$ such that $\mathsf{Sng}(x) \wedge \mathsf{Dg}(x)$. We can simulate the sort $\mathfrak{c}_k$ by arbitrary objects and we can interpret the predicate $\langle x_1, \dots, x_k \rangle \in y$ as

$$\exists z \, (z \in y \wedge \mathsf{Sng}(z) \wedge \mathsf{Tp}(z, x_1, \dots, x_k, x_1, \dots, x_1)).$$

This simulation is almost an interpretation of $\mathsf{PC}_{\leqslant n}(T)$ and the only reason why it is not (in the sense of interpretation employed in the present paper) is that we interpret different sorts by overlapping domains. However, in fact this does not matter for all the arguments in the previous parts of the paper and, by the same argument[†] as in the proof of Theorem 1, we get

**Lemma 14.** *No consistent finitely axiomatizable theory $T$ can one-dimensionally interpret $\mathsf{PC}^{\mathsf{st}}(T)$.*

Combining Lemma 14 with Lemma 13, we get

**Corollary 2.** *No consistent finitely axiomatizable $T$ can $n$-dimensionally interpret $\mathsf{PC}^{\mathsf{st}}(T^n)$.*

Since, clearly, there is a one-dimensional interpretation of the theory $\mathsf{PC}^{\mathsf{st}}(T^n)$ in the theory $\mathsf{PC}_{\leqslant n}(T)$, Corollary 2 implies Theorem 4.

## 7 | ADJUNCTIVE CLASSES MEET ADJUNCTIVE SETS

Lemma 10 is the key part of the proof of Theorem 1. In this section, we sketch a proof of a more general version of this result that might be interesting on its own.

Let $\mathsf{AC}_{\leqslant n}(T)$ be the theory in the same language as $\mathsf{PC}_{\leqslant n}(T)$. With the following axioms.

---

[†] We only would like to note that here it is slightly harder to prove the counterpart of item (3) $T \rhd_1 T \sqcup \forall x (x = x)$ from the proof of Theorem 1. Namely now we obtain this using the interpretation $T \rhd_1 \mathsf{PC}^{\mathsf{st}}(T)$ by interpreting $T$ by singletons and the additional element by the empty set.

(1) All axioms of $T$ restricted to the domain $\mathfrak{o}$.

(2) $\exists X^{\mathfrak{c}_k} \forall x_1, \ldots, x_k \neg \langle x_1, \ldots, x_k \rangle \in X^{\mathfrak{c}_k}$.

(3) $\forall X^{\mathfrak{c}_k}, x_1, \ldots, x_k \exists Y^{\mathfrak{c}_k} \forall y_1, \ldots, y_k$
$(\langle y_1, \ldots, y_k \rangle \in Y^{\mathfrak{c}_k} \leftrightarrow (\langle y_1, \ldots, y_k \rangle \in X^{\mathfrak{c}_k} \vee \bigwedge_{1 \leqslant i < k} y_i = x_i))$.

Let $\mathsf{PS}_{\leqslant n}(T)$ be the extension of $\mathsf{AC}_{\leqslant n}(T)$ be the following predicative separation scheme:

$$\forall \vec{p}, X^{\mathfrak{c}_1} \exists Y^{\mathfrak{c}_k} \forall x_1^{\mathfrak{o}}, \ldots, x_k^{\mathfrak{o}} \left( \langle x_1, \ldots, x_k \rangle \in Y \leftrightarrow \varphi(x_1, \ldots, x_k, \vec{p}) \wedge \bigwedge_{1 \leqslant i \leqslant k} x_i \in X \right),$$

where $\varphi$ is a formula such that all quantifiers in it are over the $\mathfrak{o}$-sort.

We define the no-universe axiom NU as follows:

$$\mathsf{NU} \quad \neg \exists X^{\mathfrak{c}_1} \forall x^{\mathfrak{o}} \ x \in X$$

Inspection of the part of the proof of Lemma 10 where we defined the forcing-interpretation yields the following sharper lemma.

**Lemma 15.** *There is a forcing-interpretation of* $\mathsf{AS}(T)$ *in* $\mathsf{PS}_{\leqslant 2}(T) + \mathsf{NU}$.

*Proof.* We modify the proof of Lemma 10. In the first part, we define in $\mathsf{PC}_{\leqslant n}(T)$ the notion of a small $\mathfrak{c}_1$-class and prove that the class of small $\mathfrak{c}_1$-classes is closed under adjunctions of elements. In the second part, we use this notion of smallness to define an interpretation of $\mathsf{KM}(\mathsf{AS}(T))$.

In the present case, the first part becomes superfluous and for the purpose of the second part we simply consider all classes to be small. Indeed one could see that the proof uses that small $\mathfrak{c}_1$-classes are closed under adjunction, that they satisfy no-universe axiom and we use predicative comprehension to form binary relations on a given small domain (this usage of comprehension could be replaced with the usage of separation). Specifically this properties of small sets are required for the verification of forceability of the axioms of empty set and adjunction. □

*Remark* 3. It is very well possible that there is also a non-forcing-interpretation for the same result. However, it is easy to see that we cannot generally get a non-forcing-interpretation that preserves $T$ identically on the object sort.

**Lemma 16.** *Any finite fragment of* $\mathsf{PS}_{\leqslant n}(T)$ *is interpretable in* $\mathsf{AC}_{\leqslant n}(T)$.

*Proof.* We fix a finite fragment $U$ of $\mathsf{PS}_{\leqslant n}(T)$. Suppose all the instances of the predicative separation present in $U$ are:

$$\forall \vec{p}_i, X^{\mathfrak{c}_1} \exists Y^{\mathfrak{c}_{k_i}} \forall x_1^{\mathfrak{o}}, \ldots, x_{k_i}^{\mathfrak{o}} \left( \langle x_1, \ldots, x_{k_i} \rangle \in Y \leftrightarrow (\varphi_i(x_1, \ldots, x_{k_i}, \vec{p}_i) \wedge \bigwedge_{1 \leqslant j \leqslant k_i} x_j \in X) \right),$$

for $i$ from 1 to $m$. We work in $\mathsf{AC}_{\leqslant n}$ to define the interpretation of $U$. We take the identity interpretation for the $\mathfrak{o}$-domain and the signature of $T$ as well as the interpretations of $\mathfrak{c}_k$-class domains for $k > 1$. We interpret the $\mathfrak{c}_1$-classes by restricting the domain of these classes. For the rest of the proof, we define this restriction.

We say that a $\mathfrak{c}_k$-class $X$ is *union friendly*, if for any $\mathfrak{c}_k$-class $Y$, there exists a $\mathfrak{c}_k$-class $X \cup Y$, that is, a $\mathfrak{c}_k$-class $Z$ such that

$$\forall x_1, \ldots, x_k \, (\langle x_1, \ldots, x_k \rangle \in Z \leftrightarrow (\langle x_1, \ldots, x_k \rangle \in X \vee \langle x_1, \ldots, x_k \rangle \in Y)).$$

We say that a $\mathfrak{c}_k$-class $X$ is *intersection friendly*, if for any $\mathfrak{c}_k$-class $X' \subseteq X$ and a $\mathfrak{c}_k$-class $Y$, there exists a $\mathfrak{c}_k$-class $X' \cap Y$, that is, a $\mathfrak{c}_k$-class $Z$ such that

$$\forall x_1, \ldots, x_k \, (\langle x_1, \ldots, x_k \rangle \in Z \leftrightarrow (\langle x_1, \ldots, x_k \rangle \in X' \wedge \langle x_1, \ldots, x_k \rangle \in Y)).$$

Note that a subclass of an intersection friendly class is always intersection friendly.

The domain of interpretation for $\mathfrak{c}_1$-classes consists of all intersection friendly classes $X$ such that for any $1 \leqslant i \leqslant m$, $\mathfrak{c}_1$-class $Y \subseteq X$, parameters $\vec{p}_i$, sets $B \subseteq \{1, \ldots, k_i\}$, and $\mathfrak{o}$-elements $z_1, \ldots, z_{k_i}$ there exists a union friendly class

$$Z^{\mathfrak{c}_{k_i}} = \left\{ \langle x_1, \ldots, x_{k_i} \rangle \mid \varphi_i(x_1, \ldots, x_{k_i}, \vec{p}_i), \bigwedge_{j \in \overline{B}} x_j = z_j, \text{ and } \bigwedge_{j \in B} x_j \in Y \right\}. \tag{1}$$

It is easy to see that if a $\mathfrak{c}_1$-class $X$ is in the domain of the interpretation for the sort of $\mathfrak{c}_1$-classes, then all its subclasses are in the domain of the interpretation. It is straightforward to check that all the instances of the predicative separation schemata that are among axioms of $U$ hold in this interpretation. In the case that $k_i = 1$, we use the downward closure of the domain of interpretation to see that the promised $\mathfrak{c}_1$-class $Y$ is indeed in this domain.

Clearly, the empty $\mathfrak{c}_1$-class is in the domain of interpretation. To finish the proof, we check that the adjunction axiom holds in the interpretation. So, in the rest of the proof, we check that, for any $\mathfrak{c}_1$-class $X$ from the domain of the interpretation and any $\mathfrak{o}$-element $x$, the class $X \cup \{x\}$ is in the domain of the interpretation. It is easy to see that $X \cup \{x\}$ is intersection friendly. So to finish the verification of adjunction axiom, we fix an $\mathfrak{c}_1$-class $Y \subseteq X \cup \{x\}$, $1 \leqslant i \leqslant m$, parameters $\vec{p}_i$, a set $B \subseteq \{1, \ldots, k_i\}$, and $\mathfrak{o}$-elements $z_1, \ldots, z_{k_i}$ and show that there exists a union friendly $\mathfrak{c}_k$-class

$$Z^{\mathfrak{c}_{k_i}} = \left\{ \langle x_1, \ldots, x_{k_i} \rangle \mid \varphi_i(x_1, \ldots, x_{k_i}, \vec{p}_i), \bigwedge_{j \in \overline{B}} x_j = z_j, \text{ and } \bigwedge_{j \in B} x_j \in Y \right\}.$$

If $Y \subseteq X$, then there is nothing to check, hence, in the rest of the proof, we assume that $x \in Y$. We have

$$Z^{\mathfrak{c}_{k_i}} = \bigcup_{B' \subseteq B} \left\{ \langle x_1, \ldots, x_{k_i} \rangle \mid \varphi_i(x_1, \ldots, x_{k_i}, \vec{p}_i), \right.$$

$$\left. \bigwedge_{j \in \overline{B}} x_j = z_j, \bigwedge_{j \in B'} x_j \in X \cap Y \text{ and } \bigwedge_{j \in B \setminus B'} x_j = x \right\}.$$

We observe that, since we have (1) for $X$, all individual classes in this union exist and are union friendly. This finishes the proof since, clearly, a finite union of union friendly classes is union friendly. □

Since the interpretations constructed in Lemma 16 simply restricted the $\mathfrak{c}_1$-domain, in fact they preserve the NU-axiom and thus we have

**Lemma 17.** *Any finite fragment of* $\mathsf{PS}_{\leqslant n}(T) + \mathsf{NU}$ *is interpretable* $\mathsf{AC}_{\leqslant n}(T) + \mathsf{NU}$.

**Corollary 3.** *There is a forcing-interpretation of* $\mathsf{AS}(T)$ *in* $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$.

*Proof.* By Lemma 15, we have a forcing-interpretation of $\mathsf{AS}(T)$ in $\mathsf{PS}_{\leqslant 2}(T) + \mathsf{NU}$. Inspection of the construction shows that we need certain instances of PS scheme that are required to verify the forceability of adjunction and empty set axioms. In fact, these instances do not depend on particular theory $T$. Thus $\mathsf{AS}(T)$ is interpretable in a finite fragment of $\mathsf{PS}_{\leqslant 2}(T) + \mathsf{NU}$. Hence by Lemma 17 we have an interpretation of $\mathsf{AS}(T)$ in $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$. □

**Lemma 18.** *There is an interpretation of* $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$ *in* $\mathsf{AS}(T)$.

*Proof.* It is easy to prove we can interpret $\mathsf{AC}_{\leqslant 2}$ plus the theory of a total, injective, non-surjective binary relation $\mathsf{InS}$ in $\mathsf{AS}$. See [13] for a precise definition of $\mathsf{InS}$. Then, the interpretability of $\mathsf{AC}_{\leqslant 2} + \mathsf{NU}$ follows by the result of [13, Section 5.2]. □

Combining Lemma 18 and Corollary 3, we get

**Theorem 5.** *The theories* $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$ *in* $\mathsf{AS}(T)$ *are mutually forcing-interpretable.*

Inspecting the proofs, we can see that the result is even a bit better. Both interpretations are $\mathfrak{o}$-direct and they identically translate $T$ in the $\mathfrak{o}$-sort.

**Lemma 19.** *Suppose $T$ is finitely axiomatized theory such that there is a one-dimensional interpretation of $T \sqcup \forall x\,(x = x)$ in $T$. Then, for a sufficiently large $n$, there is an interpretation of* $\mathsf{PS}_{\leqslant n}(T) + \mathsf{NU}$ *in* $\mathsf{PS}_{\leqslant n}(T)$.

*Proof.* As discussed in the proof of Lemma 15, the proof of Lemma 10 splits into two parts. The present Lemma is obtained by the first part of the proof. Namely we use the same definition of a small class in $\mathsf{PS}_{\leqslant n}(T)$, although now we do not know whether there exists the $\mathfrak{c}_1$-class of all elements. None the less, the same proof as in Lemma 10 shows that if it exists, then it is not small. Also the same proof as before shows that small $\mathfrak{c}_1$-classes are closed under adjunctions. Thus, we can interpret $\mathsf{PS}_{\leqslant n}(T) + \mathsf{NU}$ in $\mathsf{PS}_{\leqslant n}(T)$ by keeping everything as is, but restricting the domain of $\mathfrak{c}_1$-classes to small $\mathfrak{c}_1$-classes. □

**Corollary 4.** *Suppose $T$ is a finitely axiomatized theory that one-dimensionally interprets $T \sqcup \forall x(x = x)$. Then, for sufficiently large $n$, the theory* $\mathsf{AC}_{\leqslant n}(T)$ *forcing-interprets* $\mathsf{AS}(T)$.

*Proof.* Since $T$ is finitely axiomatized, the theory $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$ is also finitely axiomatized and, hence, by Lemma 19, the theory $\mathsf{AC}_{\leqslant 2}(T) + \mathsf{NU}$ is interpretable in $\mathsf{PS}_{\leqslant n}(T)$, for some $n$. Since the

theory $AC_{\leqslant 2}(T) + NU$ is finitely axiomatized, it is interpretable in a finite fragment of $PS_{\leqslant n}(T)$ and, by Lemma 16, in $AC_{\leqslant n}(T)$. By Lemmas 15 and 8, we get a forcing-interpretation of $AS(T)$ in $AC_{\leqslant n}(T)$. $\qquad\square$

**Corollary 5.** *Suppose finitely axiomatizable* $T \rhd_1 AC(T)$. *Then* $T$ *forcing-interprets* $AS(T)$.

*Proof.* Clearly $AC(T)$ interprets $T \sqcup \forall x(x = x)$. Note that, if we replace $PC_{\leqslant n}$ with $AC_{\leqslant n}$ in all the lemmas from Section 3, all the proofs work without any modifications. In particular, by the modified version of Lemma 7, for each $n$, the theory $AC_{\leqslant 2^n}(T)$ is interpretable in $AC^{2n+1}(T)$. Thus, for each $n$, the theory $AC_{\leqslant n}(T)$ is intepretable in $T$. Hence, by Corollary 4 and Lemma 8, the theory $T$ forcing-interprets $AS(T)$. $\qquad\square$

## 8 | QUESTIONS AND PERSPECTIVES

Our paper points to several potential directions of further research.

Despite the fact that the formulation of Theorem 1 does not employ arithmetization, the proof reduces the result to the usual Gödel's Second Incompleteness Theorem. Hence we have the following question.

(1) Find a more direct proof of Theorem 1 that does not employ arithmetization.

There are questions about generalizing Theorems 1 and 4:

(2) Is there a finitely axiomatizable theory $T$ without finite models that does interpret $PC(T)$?
(3) Is there a theory $T$ axiomatized by finitely many schemes that one-dimensionally interprets $PC^{schem}(T)$?
(4) Is there a finitely axiomatizable theory $T$ without finite models that one-dimensionally interprets $KM(PC(T))$?

A downside of the main result of this paper is that it does not establish PC as a jump operator, since our result is applicable to finitely axiomatizable theories, but in general we do not have reasons to believe that $PC(T)$ is finitely axiomatizable for all finitely axiomatizable theories $T$. Thus we have the following question.

(5) Is it true that for any finitely axiomatized $T$ there is a finitely axiomatizable subtheory $T'$ of $PC(T)$ such that $T$ does not one-dimensionally interpret $T'$?

Ideally, the theories $T'$ should be defined by some natural and uniform construction from $T$.

There are questions about the behaviour of PC operator on (interpretability) weak theories.

(6) Characterize the interpretability degree of $PC(T)$ for classical decidable theories like $Th(\mathbb{N}, +)$, $Th(\mathbb{N}, \times)$, $Th(\mathbb{N}, S)$, $Th(\mathbb{N}, <)$, $Th(\mathbb{Q}, <)$, $Th(\mathbb{R}, 0, 1, +, \times)$, $Th(\mathbb{R}, 0, +)$.

Also, it might be interesting to figure out the interaction of PC operator with various tameness notions from model theory.

Basic facts about forcing-interpretations need to be developed. We need things like a precise definition of composition and the verification of its desired properties. An attractive way to do that would be to view the category of forcing-interpretations as a co-Kleisli category. The ingredients for the desired co-monad KM would be the identical one-world interpretation from $KM(T)$ in $T$ and

an interpretation of $KM(T)$ in $KM(KM(T))$, where worlds are interpreted as pairs of worlds. A further issue is sameness of forcing-interpretations and the related question about the 2-category of forcing-interpretations. We can simply take over notions of sameness/isomorphism from ordinary interpretations, but we can also think of new ones, for example, ones inspired by bisimulations of Kripke models.

The central part of our argument is the forcing-interpretation of adjunctive set-theory. So there is a natural question, if forcing was necessary here.

(7) Is there an interpretation of $AS(T)$ in $PS_{\leqslant 2}(T) + NU$, for finitely axiomatizable theories?

(8) Is there always an interpretation of $AS(T)$ in $PS_{\leqslant 2}(T) + NU$?

(9) Generally, in which circumstances can forcing-interpretations be replaced with interpretations? In the case of finitely axiomatized sequential theories or reflexive sequential theories, there is an argument that this can be done. However, even for arbitrary sequential theories, we do not know whether this is always possible.

## JOURNAL INFORMATION

The *Bulletin of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

## REFERENCES

1. J. Avigad, *Eliminating definitions and Skolem functions in first-order logic*, ACM Trans. Comput. Log. (TOCL) **4** (2003), no. 3, 402–415.
2. J. Avigad, *Forcing in proof theory*, Bull. Symb. Log. **10** (2004), no. 3, 305–333.
3. L. D. Beklemishev, *Gödel incompleteness theorems and the limits of their applicability. I*, Russian Math. Surveys **65** (2010), no. 5, 857.
4. S. R. Buss, *Bounded Arithmetic*, Bibliopolis, Napoli, 1986.
5. S. Feferman, *Arithmetization of metamathematics in a general setting*, Fund. Math. **49** (1960), 35–92.
6. J. Ferrante and C. W. Rackoff, *The computational complexity of logical theories*. Springer, Berlin, 1979.
7. T. Franzén, *Gödel's theorem: an incomplete guide to its use and abuse*. AK Peters/CRC Press, Boca Raton, F.L., 2005.
8. K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für mathematik und physik **38** (1931), no. 1, 173–198.
9. P. Hájek and P. Pudlák, *Metamathematics of first-order arithmetic*, vol. 3. Cambridge Univ. Press, Cambridge, 2017.
10. S. G. Simpson and S. G. Simpson, *Subsystems of second order arithmetic*, vol. 1. Cambridge Univ. Press, Cambridge, 2009.
11. C. Smoryński, *The Incompleteness Theorems*. In J. Barwise (ed.), Handbook of Mathematical Logic, North-Holland, Amsterdam, 1977, pp. 821–865.
12. A. S. Troelstra and D. van Dalen, *Constructivism in Mathematics, vol. 1*, Stud. Logic Found. Math., vol. 121, North Holland, Amsterdam, 1988.
13. A. Visser, *Cardinal arithmetic in the style of Baron von Münchhausen*, Rev. Symb. Log. **2** (2009), no. 3, 570–589.
14. A. Visser, *The predicative Frege hierarchy*, Ann. Pure Appl. Logic **160** (2009), no. 2, 129–153.

15. A. Visser, *Can we make the Second Incompleteness Theorem coordinate free?* J. Logic Comput. **21** (2011), no. 4, 543–560.
16. A. Visser, *On* Q, Soft Comput. **21** (2017), no. 1, 39–56.
17. A. Visser, *The small-is-very-small principle*, Math. Log. Q. **65** (2019), no. 4, 453–478.