THE BULLETIN OF SYMBOLIC LOGIC
Volume 19, Number 3, Sept. 2013

INTERPRETABILITY IN ROBINSON'S Q

FERNANDO FERREIRA AND GILDA FERREIRA

Abstract. Edward Nelson published in 1986 a book defending an extreme formalist view of mathematics according to which there is an *impassable barrier* in the totality of exponentiation. On the positive side, Nelson embarks on a program of investigating how much mathematics can be interpreted in Raphael Robinson's theory of arithmetic Q. In the shadow of this program, some very nice logical investigations and results were produced by a number of people, not only regarding what can be interpreted in Q but also what cannot be so interpreted. We explain some of these results and rely on them to discuss Nelson's position.

§1. Introduction. Let $\mathcal{L}$ be the first-order language with equality whose nonlogical symbols are the constant 0, the unary function symbol S (for successor) and two binary function symbols + (for addition) and $\cdot$ (for multiplication). The following theory was introduced in [35] (see also the systematic [42]):

DEFINITION 1. *Raphael Robinson's theory* Q *is the theory in the language* $\mathcal{L}$ *given by the following axioms*:

1. $\mathrm{S}x \neq 0$,
2. $\mathrm{S}x = \mathrm{S}y \rightarrow x = y$,
3. $x + 0 = x$,
4. $x + \mathrm{S}y = \mathrm{S}(x + y)$,
5. $x \cdot 0 = 0$,
6. $x \cdot \mathrm{S}y = x \cdot y + x$,
7. $x \neq 0 \rightarrow \exists y\, (\mathrm{S}y = x)$.

This theory is totally inadequate for the formalization of arithmetic. The reason for this lies in the fact that induction is conspicuous by its absence. The theory Q can be considered a most extreme case of a theory of weak

Received July 30, 2012.
Both authors acknowledge support of FCT–Fundação para a Ciência e a Tecnologia [PEst-OE/MAT/UI0209/2011 and PTDC/MAT/104716/2008]. The second author is also grateful to FCT [grant SFRH/BPD/34527/2006] and Núcleo de Investigação em Matemática (Universidade Lusófona).

1079-8986/13/1903-0001/$3.90
DOI:10.2178/bsl.1903010

arithmetic, one in which there is no induction.[1] Of course, Robinson's purpose in defining Q was not to formalize arithmetic in it. It was rather to present a finitely axiomatizable theory which is essentially undecidable. In 1986, Edward Nelson wrote a very interesting book, entitled "Predicative Arithmetic" [28], in which he advances an extreme formalist position which casts doubt about the ordinary mathematician's belief in the set of natural numbers.[2] Nelson's position detects an impassable barrier in the totality of exponentiation (the basic informal argument says, roughly, that the number of steps needed to terminate a recursion defining exponentiation is of the order of magnitude of exponentiation itself—a perceived circularity) and recoils into the "safe theory" Q. Indeed, Nelson does not believe in the totality of exponentiation and even suggests, in chapter 31 of his book, that one should try to develop mathematics under the assumption of the denial of the totality of exponentiation.

In the first part of his book, Nelson embarks on the project of seing how much arithmetic can be interpreted in Q. Some very nice and unexpected mathematics ensued in the shadow of Nelson's interpretability program (and related studies), not only by Nelson himself but also by Robert Solovay, Petr Hájek, Samuel Buss, Alex Wilkie, Jeff Paris, Pavel Pudlák, Albert Visser and others. In a sense, Nelson's position concerning the impassable barrier is vindicated by the twin results of Wilkie (according to which the totality of exponentiation *is not* interpretable in Q) and Solovay (the *negation* of the totality of exponentiation *is* interpretable in Q). Both these results are proved in Section 8 below. A blunt rejection of Nelson's ideas is an option—albeit not the most interesting one. It is more interesting to discuss some tensions in Nelson's position, as well as the claim that it is a *predicative* position in the foundations of mathematics. The next section includes some discussions of these issues. They are informed by technical results, and these form the main body of this article (sections 3 to 8). In the remainder of this introduction, we briefly discuss the notion of interpretability, with a special emphasis towards results of interpretability in Q.

The notion of interpretability of a theory into another was introduced by Alfred Tarski, Andrzej Mostowski and Raphael Robinson in the above cited monograph [42]. Since then, the notion has been much extended. In this introduction we adopt a non-formal key. In its most basic form, an interpretation of a theory S into a theory T consists of the following data:

---

[1] A weak theory of arithmetic is a subtheory of Peano Arithmetic in which the induction scheme is restricted. When the restriction is to bounded formulas (see Section 4), one usually speaks of bounded theories of arithmetic.

[2] The book has some tantalizing thoughts. E.g., "The mathematician's attitude towards ω has in practice been one of faith, and faith in a hypothetical entity of our own devising, to which are ascribed attributes of necessary existence and infinite magnitude, is idolatry" (in chapter 18).

(a) a one-place formula in the language of T, the *domain formula*, which is meant to be the (non-empty) domain of objects of S from the point of view of T;

(b) appropriate definitions of constants, function symbols and relation symbols of the language of S in the language of T;

together with the requirement that

(c) every axiom of S, when translated into the language of T (based on the previous clauses), is a theorem of T.

The translation referred to above is the well-known translation that commutes with the propositional connectives. Also, as it is usual, if $J(x)$ is the domain formula, then the translation of $\forall x\, \varphi$ is $\forall x\, (J(x) \to \varphi^T)$ and the translation of $\exists x\, \varphi$ is $\exists x\, (J(x) \wedge \varphi^T)$, where φ^T is the translation of the formula φ. It is handy to use the notations $\forall x^J\, \varphi^T$ and $\exists x^J\, \varphi^T$ for the second and fourth formulas of the previous sentence, as well as for similar situations. Note that the notion of interpretability is transitive. This fact will be used throughout without ado.

In theories of arithmetic which are strong enough to have the "less than or equal" relation $\leq$ with its main properties (e.g., the theory Q^+ in Section 3), a very important form of interpretability is *cut-interpretability*. In this case, the domain formula is provably downwards closed with respect to $\leq$ and the constant and function symbols are interpreted by themselves (see Section 3 for details). Notice that if T and S are theories in the language of arithmetic, and the former is true, and S is cut-interpretable in T, then S must also be a true theory. To see this, just observe that since the standard structure $\mathbb{N}$ is a model of T and the domain formula is closed under successor (and has 0), then $\mathbb{N}$ is also a model of S. Many interpretations of this paper are cut-interpretations (e.g., all the interpretations into Q^+ in sections 3 to 6). However, Solovay's interpretation mentioned earlier cannot be a cut-interpretation, since it interprets a falsity (the denial of the totality of exponentiation).

The above notion of interpretability can be relaxed. For instance, parameters may be allowed in the definitions. The interpretation may also be piecewise. This feature shows up in the interpretations of theories of analysis discussed in Section 7. These theories are two-sorted, with a sort for numbers and another sort for sets of numbers. Each sort is interpreted by a formula, defining a *piece*, and the pieces may be overlapping (an element of the interpreting theory may play a double role, one of standing for an element of the number sort, the other in the role of standing for an element of the set sort). It is not necessary to give—if it is at all possible to give—the most general definition of interpretability. We only require (in order to prove non-interpretability results) that our notion of interpretation satisfies a very basic property. In a nutshell, we assume that if a theory S is interpretable in a theory T then the consistency of the former follows finitistically from the consistency of the latter. This is certainly true of all forms of interpretation in this article. Furthermore, we require that this relative consistency property is provable in a certain weak theory of arithmetic. Let us be more precise. We work with a theory of arithmetic in which the arithmetization of syntax poses no problems. This is the case with the theory of bounded arithmetic S^1_2, introduced by Samuel Buss in [5]. In this theory, the basic syntactic notions and operations can be defined and their basic properties can be proved. It is possible to define standard Hilbert-style consistency predicates 'Con(T)' for finitely axiomatizable theories T, and Gödel's incompleteness theorems can be proved (we discuss the theory S^1_2 in sections 4 and 5). We assume that our notion of interpretability enjoys the following property: If S and T are finitely axiomatizable theories and the former is interpretable in the latter, then $\mathsf{S}^1_2 \vdash \mathrm{Con}(T) \to \mathrm{Con}(S)$.

The four sections 3–6 are concerned with interpretability of theories of arithmetic into Q. Some classical and beautiful results are presented. Section 7 concerns the interpretability of theories of analysis (second-order theories) into Q. It was shown in [15] that a theory of "feasible analysis" BTFA (an acronym for Base Theory for Feasible Analysis) is interpretable in Q. In this paper, we extend this result to a theory of analysis BTPSA related to the polyspace computable functions. The theory BTFA already formalizes and proves the very basic properties of real numbers and continuous functions, being able to deduce the intermediate value theorem. As a consequence, Tarski's well-known theory of real closed ordered fields RCOF ([41]; see [27] for a modern reference) is interpretable in BTFA. We get a very satisfying picture: the theory RCOF (the archetypal "tame" theory) is interpretable into Q (the archetypal "untame" theory), but not the other way around (because Q is essentially undecidable and RCOF is decidable).[3,4] It is also worth pointing out that (first-order) Euclidean geometry can be formalized in RCOF and that, therefore, Euclidean geometry is interpretable in Q (see section 3 of [23] for more information on related metamathematical results concerning Euclidean geometry).

The interpretability of BTPSA into Q is not surprising (it relies on an adaptation of the argument for BTFA). However, it should be mentioned that BTPSA is strong enough to develop Riemann integration up to the fundamental theorem of calculus. It is really a bit surprising how much mathematics can be interpreted in Q.

We are grateful to Robert Solovay for the kind permission to report on his old unpublished results obtained in the mid ninety eighties (see Section 8). We also want to thank the anonymous referee for many suggestions of improvement.

[3]Harvey Friedman has also announced in [22] the interpretability of RCOF in Q.

[4]The distinction between tame and untame theories is intuitive. Roughly, an untame theory is one subjected to Gödel's incompleteness phenomenon, while tame theories are, as a rule, decidable and of a more algebraic or geometric nature.

§**2. A form of predicativity?** In its most basic form, predicativism is a position concerning the definitions of sets. A specification of a set is *impredicative* if it generalizes over a totality of sets over which the specified set belongs. Predicativism only accepts sets insofar as they are given by definitions and, therefore, detects a vicious circle in impredicative specifications. Sets, according to the predicativist, do not have an independent existence: they are created by man (via definitions). Traditional predicativism in the foundations of mathematics accepts the natural numbers as given and adopts a predicative position regarding sets of natural numbers. This classical view has its roots in [47] and was the object of much attention by proof-theorists in the sixties (the *locus classicus* of this discussion is [11]).

The appearance of the notion of impredicativity in the debate on foundations of mathematics is due to Henri Poincaré in [33] (after a seminal observation of Jules Richard). This notion surfaced because of discussions around the Frege–Dedekind definition of the set of natural numbers. In these debates, the set of natural numbers is not given, it is rather defined within an infinite Dedekind domain (a domain that satisfies axioms 1 and 2 of Q): the set of natural numbers is defined as the least inductive set (i.e., the least set with 0 and closed under successor). The principle of induction is supposed to follow from the definition. This definition is impredicative as noticed by Poincaré. Observe that we are no longer discussing predicativity *given* the natural numbers but are questioning the predicativity of the natural number notion itself. This kind of predicativity, which is conceptually different from the more traditional predicativism that accepts the natural numbers as given, was coined *strict predicativity* by Charles Parsons in [32].

In a strict predicative treatment of arithmetic, one must always be on guard against unpermissible forms of proof by induction. Since the natural numbers are defined by the formula

$$\mathbb{N}(x) :\equiv \forall X\,(0 \in X \wedge \forall x\,(x \in X \rightarrow \mathrm{S}x \in X) \rightarrow x \in X),$$

and because we are adopting a predicative position regarding sets, this means that we only have induction for predicatively defined sets X. The problem lies in the fact that the natural numbers do not form a set (they are defined impredicatively) and, therefore, it is not permitted to argue by induction with respect to formulas in which the notion of natural number occurs. This restriction on induction seems to seriously paralyze the development of arithmetic. Indeed, how much arithmetic can be developed in a strict predicative manner?

The central technical paper concerning strict predicativity is due to John Burgess and Allen Hazen [4]. They worked with a so-called *ramified* predicative system. Second-order quantifications are assigned a specific level (0, 1, 2, etc.). A set of level k can be formed as long as it is the extension of a formula in which the occurrences of second-order quantifications are of level strictly less than k (first-order parameters and set parameters up to, and including, level k are allowed). The ramification of second-order quantification into levels is predicatively acceptable. Within this ramified setting, there is a natural number notion corresponding to each non-zero level. Burgess and Hazen showed that level 2 numbers form a model of bounded arithmetic with exponentiation. On the strict predicative basis that we are discussing, the impassable barrier of Nelson is overcome.

Nelson's position is *prima facie* predicative since it starts with Q, where no induction is present, and seems to proceed predicatively thereafter. However, this first evaluation is careless. There are serious problems with merely accepting those sentences which are interpretable in Q. One of these problems is already mentioned by Nelson himself in chapter 15 of his book [28]. It is the so-called *compatibility problem*. Is it the case that if ϕ_0 and ϕ_1 are interpretable in Q then their conjunction $\phi_0 \wedge \phi_1$ is also interpretable in Q? The answer is negative, and in the proof of Proposition 8 we present an argument of Solovay to that effect. Given this fact, the sensible position seems to be to accept not only the sentences which are interpretable in Q but (of course) also their logical consequences. This natural extension does not work, though. Solovay also showed that there is a sentence ϕ which is interpretable in Q, as well as its *negation* (see Proposition 9). A possible way out of this conundrum is to accept the sentences which are cut-interpretable in Q^+ (and their logical consequences). Since sentences cut-interpretable in Q^+ are true, we know—at least from an impredicative viewpoint—that these sentences do cohere. However, the very same sentences ϕ_0 and ϕ_1 mentioned above (see the proof of Proposition 8) are cut-interpretable in Q^+ but their conjunction entails the totality of exponentiation. This position goes over the impassable barrier. It is perhaps interesting to investigate how much does it go over and what is its relation with strict predicativity, but we would certainly be departing from Nelson's position—one which takes the impassability of the barrier of exponentiation as its central tenet.

Nelson's predicativism suffers from a certain instability. At the end of chapter 15 of his book, Nelson says that "it is necessary to take care (...) that our methods for adding a formula to the growing list of predicatively established formulas do not interfere with any previously established formula." What directs this care? Given incompatible interpretations, how does one choose between them? How are we sure that, even preserving consistency, we will not pass the impassable? What is the principled reason that directs the choice of interpretable sentences?

There is another angle from which Nelson's position can be commented upon. It concerns the reasons why it does not accept second-order predicative reasoning. It seems to us that there is at play in Nelson's position some ingredients that go beyond mere predicativism (at least as ordinarily conceived). Nelson himself (e.g., [30]) has suggested that there are connections between his brand of predicative arithmetic and the work of Bellantoni and Cook [2] on tiered recursion. Perhaps this work can explain why exponentiation is impredicative in a way that weaker recursions are not. However, the results of Bellantoni and Cook concern *definitions* of functions by recursion whereas Nelson's predicativism concerns *theories* of arithmetic. For all the suggestions of a cogent relationship between the two approaches there is not, as far as we are aware, a detailed and sustained study connecting the two positions which, for instance, tries to respond—in a principled way—to the questions of the previous paragraph.

We do not want to leave this section without referring to an important factor that is also at play in Nelson's position, one that is not directly connected with predicativity. This factor is Nelson's extreme formalism (see [29] for a particularly clear apology of this position). We classify Nelson's formalism as extreme because it does not rest on finitism, as does the formalism of Hilbert. It rests on a much more strict base, one connected with his predicativism and, in particular, one which does not accept the totality of exponentiation. The extreme formalism of Nelson also shows up in his manifested doubts about the consistency of Peano Arithmetic. In his most recent paper [31], Nelson writes that "perhaps [Peano Arithmetic] is inconsistent."[5] One may add that Nelson is not alone among first-rate mathematicians who have expressed doubts about the consistency of Peano Arithmetic: Jacob T. Schwartz in [37] and Vladimir Voevodsky in [46] have expressed similar doubts. We will not pursue this matter further.

§3. The local interpretability of $\mathsf{I}\Delta_0$ in Q. In the introduction, we said that the theory Q is inadequate to formalize arithmetic. For instance, Q does not even prove that addition is associative. To show this, let us briefly describe an example of [42]. Construct an interpretation of Q in the following way. Its domain is $U :\equiv \mathbb{N} \cup \{\infty_0, \infty_1\}$, where ∞_0 and ∞_1 are distinct elements different from the natural numbers. 0 has its usual meaning, as well as the arithmetical operations when applied to the natural numbers. For $i = 0, 1$, we let $\mathrm{S}(\infty_i) = \infty_i$; $\infty_i + n = \infty_i$, for $n \in \mathbb{N}$; $u + \infty_i = \infty_{1-i}$, for $u \in U$; $n \cdot \infty_i = \infty_i$, for $n \in \mathbb{N}$; $\infty_i \cdot 0 = 0$; and $\infty_i \cdot u = \infty_{1-i}$, for $u \in U \setminus \{0\}$. It is easily checked that this forms a model of Q. However, $(0 + \infty_0) + \infty_0 = \infty_1$ and $0 + (\infty_0 + \infty_0) = \infty_0$. Note that $\forall x\,(\mathrm{S}x \neq x)$, $\forall x\,(0 + x = x)$ and $\forall x\,(0 \cdot x = 0)$ also fail in this model. It is perhaps worth remarking that this model (and similar ones) can be easily converted into an interpretation of Q together with the failure of (say) associativity of addition into Q.[6]

[5] He wrote in his book (end of chapter 31 of [28]) that he has "put a lot of effort into this [in proving in Peano Arithmetic the negation of the totality of exponentiation, a result which would, of course, entail the inconsistency of Peano Arithmetic], but *so far* without success" (our italics).

[6] Just shift the numerical domain with $+2$ making $\infty_0 := 0$ and $\infty_1 := 1$, induce the new arithmetical operations on the new numbers using the old ones shifting back and forth, and

It was also mentioned in the introduction that it is nevertheless a bit surprising how much can be interpreted in Q. The purpose of this section is to present some basic techniques of interpretability in Q. To get "off the ground" in doing this, the following theory and the next theorem are important. Let Q^+ be the extension of Q with the following extra axioms:

8. $(x + y) + z = x + (y + z)$,
9. $x \cdot (y + z) = x \cdot y + x \cdot z$,
10. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$,
11. $x + y = y + x$,
12. $x \cdot y = y \cdot x$,
13. $x \leq y \leftrightarrow \exists z\,(x + z = y)$.

In the above, the language $\mathcal{L}$ was also extended, say to the language $\mathcal{L}^+$ which includes the binary relation symbol $\leq$.

THEOREM 1. *Q^+ is interpretable in Q.*[7]

The proof of this result is *ad hoc* and takes a bit over six pages in chapter 6 of Nelson's monograph [28]. Actually, Nelson works with a reformulation Q^P of Q. This new theory adds to the language $\mathcal{L}$ a unary function symbol P (for predecessor) and replaces axiom 7 by

7^P. $Px = y \leftrightarrow \mathrm{S}y = x \vee (x = 0 \wedge y = 0)$.

It is clear that the theory Q^P is interpretable in Q (the former is an extension by definitions of the latter) and vice versa. The theory Q^P is occasionally useful on account of being an open theory. The option of starting with Q instead of the more suave theory Q^+ can be explained by historical reasons (Q is the traditional finitely axiomatizable, essentially undecidable, theory) and "for the pleasure of working from minimal assumptions" (quoted from chapter 6 of [28]). Nothing important is lost if one starts with Q^+ instead. With theories that include Q^+, we can be more systematic.

DEFINITION 2. *Let T be a theory that contains Q^+. A formula $\varphi(x)$ is* inductive *in T if*

$$T \vdash \varphi(0) \wedge \forall x\,(\varphi(x) \rightarrow \varphi(\mathrm{S}x)).$$

DEFINITION 3. *Let T be a theory that contains Q^+. A formula J is a* cut *in T if it is inductive and satisfies*

$$T \vdash J(x) \wedge y \leq x \rightarrow J(y).$$

A formula $J(x)$ is a subcut *of the formula $I(x)$ in T if, moreover,*

$$T \vdash J(x) \rightarrow I(x).$$

define the operations involving the ∞'s "by hand." We thank the anonymous referee for pointing us this kind of conversion.

[7]As a consequence of this theorem and the discussion in the first paragraph of this section, we obtain examples of *Orey sentences* ϕ for Q, i.e., of sentences ϕ such that both $\mathsf{Q} + \phi$ and $\mathsf{Q} + \neg\phi$ are interpretable in Q. In Proposition 9, we provide a more interesting Orey sentence.

Lemma 1. *Let T be a theory that contains* Q^+ *and consider $I(x)$ an inductive formula in T. Then there exists a subcut $J(x)$ of $I(x)$ in T.*

Proof. Fix $I(x)$ an inductive formula in Q^+. Define

$$J(x) :\equiv \forall y\,(y \leq x \rightarrow I(y)).$$

Let us prove that J is inductive. In Q^+ it can be proved that $y \leq 0 \rightarrow y = 0$, so $J(0)$ is equivalent to $I(0)$. Therefore, we get $J(0)$. Assume $J(x)$ and let $y \leq \mathrm{S}x$. We need to prove that we have $I(y)$. If $y = 0$, this is immediate. If $y \neq 0$, axiom 7 ensures that $y = \mathrm{S}y'$ for some y'. Since Q^+ proves the equivalence $x \leq y \leftrightarrow \mathrm{S}x \leq \mathrm{S}y$, we get $y' \leq x$. By $J(x)$, we have $I(y')$. By the inductiveness of I, we conclude that $I(\mathrm{S}y')$, i.e., $I(y)$. Therefore, J is inductive.

Let us finally check that $J(x) \wedge y \leq x \rightarrow J(y)$. Suppose $J(x) \wedge y \leq x$. Let $z \leq y$. We need to show that $I(z)$. Using the transitive property of $\leq$ (easily proved in Q^+), we have $y \leq x \wedge z \leq y \rightarrow z \leq x$. Using $J(x)$, we obtain $I(z)$. $\dashv$

The following result (due independently to Solovay [40] and Nelson [28]) is fundamental:

Proposition 1. *Let T be a theory that contains* Q^+ *and consider $I(x)$ an inductive formula in T. Then there exists a cut $J(x)$ such that J is a subcut of I closed under addition and multiplication.*

Proof. By Lemma 1, we may assume that I is a cut. In order to find a subcut of I closed under the desired operations, we are going to use the technique of *shortening the cut* (due to unpublished work of Solovay). This strategy allow us to replace any cut by a cut contained in it and with some additional properties: in our case closed under addition and multiplication.

Take $J_0(x) :\equiv \forall y\,(I(y) \rightarrow I(y+x))$ and $J(x) :\equiv \forall y\,(J_0(y) \rightarrow J_0(y \cdot x))$. The fact that J is contained in J_0 results from $J_0(\mathrm{S}0)$. Since it is immediate that J_0 is contained in I, we conclude that J is contained in I. It is easy to show that J is closed under addition once we prove that J_0 is closed under addition (just use the distributive law). The mentioned closedness property of J_0 follows easily from the associativity of $+$. Now, by axiom 6 and $J(\mathrm{S}0)$, we get that J is inductive. Using the associativity for multiplication, we can similarly prove that J is closed under multiplication.

In order to see that J is a cut, it remains to show that $J(x) \wedge y \leq x$ entails $J(y)$. Fix z such that $J_0(z)$. We want to show that $J_0(zy)$, where (as usual) zy abbreviates $z \cdot y$. So, fix r such that $I(r)$ and let us prove that we have $I(r + zy)$. By hypothesis and $J_0(z)$ we have $J_0(zx)$. Since $y \leq x$, there exists w such that $y + w = x$. So, we have $J_0(z(y + w))$, i.e., $J_0(zy + zw)$. From the latter, knowing that we have $I(r)$, we get $I(r + (zy + zw))$. Since I is a cut and $r + zy \leq (r + zy) + zw = r + (zy + zw)$ we conclude the proof. $\dashv$

Let us introduce as abbreviations *bounded* quantifiers, writing

$$\forall x \leq t\,(\dots) \qquad \text{for} \qquad \forall x\,(x \leq t \rightarrow \dots),$$
$$\exists x \leq t\,(\dots) \qquad \text{for} \qquad \exists x\,(x \leq t \wedge \dots)$$

where t is a term in which the variable x does not occur.

DEFINITION 4. *A* bounded formula *is a formula which is built from atomic formulas using only propositional connectives and bounded quantifiers.*

Bounded formulas are absolute with respect to cuts which are closed under addition and multiplication. That is, if J is a cut (closed under addition and multiplication) in a theory T which extends Q^+ and $\varphi(\vec{x})$ is a bounded formula (with its free variables as shown), then $T \vdash \forall \vec{x}^J\,(\varphi^J(\vec{x}) \leftrightarrow \varphi(\vec{x}))$. This is easily checked by induction of the complexity of the (bounded) formula φ. Note that, as a consequence, if $T \vdash \forall \vec{x}\,\varphi(\vec{x})$ then $T \vdash \forall \vec{x}^J\,\varphi^J(\vec{x})$. Since axioms 7 and 13 are equivalent in Q^+ to universal closures of bounded formulas, the cut J in Proposition 1 is a model of Q^+.

DEFINITION 5. *The theory* $\mathsf{I}\Delta_0$ *is the extension of* Q^+ *with the induction scheme for the bounded formulas, i.e.,* $\varphi(0) \wedge \forall x\,(\varphi(x) \rightarrow \varphi(\mathrm{S}x)) \rightarrow \forall x\,\varphi(x)$, *where* φ *is a bounded formula* (*possibly with parameters*).

It is worth remarking that the instances of bounded induction can be written as follows:

$$\forall x\,[\varphi(0) \wedge \forall y \leq x\,(\varphi(y) \rightarrow \varphi(\mathrm{S}y)) \rightarrow \varphi(x)],$$

with φ a bounded formula. As a consequence, if a theory is a model of $\mathsf{I}\Delta_0$ then so is any of its cuts (closed under addition and multiplication). With the goal of proving the interpretability of $\mathsf{I}\Delta_0$ in Q, we first prove a weaker result.

DEFINITION 6. *A theory T is* locally interpretable *in a theory S if each finite part of T is interpretable in S.*

Note that the previous notion of interpretability is potentially weaker than (global) interpretability since, by definition, different portions of T may have different interpretations in S.[8]

The proof of the following result is easily available in the literature (see chapter 7 of [28] or [25], page 369). We include it here for the sake of completeness.

THEOREM 2. $\mathsf{I}\Delta_0$ *is locally interpretable in* Q.

PROOF. By Theorem 1 it is enough to show that $\mathsf{I}\Delta_0$ is locally interpretable in Q^+. By the definition of local interpretability, we just need to fix an arbitrary finite set $\varphi_1(x, \vec{p}), \dots, \varphi_n(x, \vec{p})$ of bounded formulas (with a vector of parameters $\vec{p}$) and prove that Q^+ plus induction for each of the $\varphi_1, \dots \varphi_n$ is interpretable in Q^+.

[8] In [45] the reader can find separating examples.

Let

$$I(x) :\equiv \forall \vec{p}\,(I_1(x,\vec{p}) \wedge \cdots \wedge I_n(x,\vec{p})),$$

where each I_i $(i = 1, \ldots, n)$ is defined by:

$$I_i(x,\vec{p}) :\equiv \varphi_i(0,\vec{p}) \wedge \forall y \leq x\,(\varphi_i(y,\vec{p}) \rightarrow \varphi_i(\mathrm{S}y,\vec{p})) \rightarrow \varphi_i(x,\vec{p}).$$

Note that each formula $I_i(x,\vec{p})$ is bounded. It is also clear that, for each i, the formula $\forall \vec{p}\, I_i(x,\vec{p})$ is inductive in Q^+. It readily follows that the formula $I(x)$ is also inductive in Q^+. By Proposition 1, we know that there is a cut J contained in I that is closed under addition and multiplication (hence, J is a model of Q^+). Since $J(x) \rightarrow I(x)$ and $I(x) \rightarrow I_i(x,\vec{p})$, for every i, induction for every φ_i holds in J. ⊣

§**4. Bounded theories.** In his great incompleteness paper, Gödel showed how to formalize syntax in a formal theory. In modern treatments of incompleteness, syntax is formalized in the theory $\mathsf{I}\Sigma_1$, the restriction of Peano Arithmetic to induction for Σ_1-formulas. Following Gödel, primitive recursive functions can be introduced (in an appropriate manner) in $\mathsf{I}\Sigma_1$. It is also possible to formalize syntax in much weaker theories. A particularly good theory for this is Buss's theory S^1_2 (briefly discussed below): in this theory, some amount of recursion is formalizable (polynomial bounded recursion on notation) and, in fact, it is possible to introduce (in an appropriate manner) the polytime computable functions in S^1_2. The usual syntactical notions are of polytime character and, therefore, S^1_2 is a weak theory enjoying a good formalization of the syntax. On the other hand, the theory $\mathsf{I}\Delta_0$ is not very convenient for the formalization of syntax because polytime functions cannot be introduced in it. The reason lies in the fact that, given two numbers x and y, one cannot find in $\mathsf{I}\Delta_0$ a number z whose binary length is the product of the binary lengths of x and y (this is easily fixed with the introduction of the axiom Ω_1: see below and, for extensive discussions, consult [5] and [49]). However, even though $\mathsf{I}\Delta_0$ does not have the resources to formalize simple forms of recursion, the following is true:

PROPOSITION 2. *There is a bounded formula* $\mathrm{Exp}(x,y,z)$, *with the free variables as shown, such that* $\mathsf{I}\Delta_0$ *proves*:

(i) $\mathrm{Exp}(x,0,z) \leftrightarrow z = 1$,
(ii) $\mathrm{Exp}(x,\mathrm{S}y,z) \leftrightarrow \exists w\,(\mathrm{Exp}(x,y,w) \wedge z = wx)$.

The above result says that the exponentiation function $x, y \rightsquigarrow x^y$ has a graph defined by a bounded formula (a result essentially due to James Bennett [3]), and that its recursive equations are provable in $\mathsf{I}\Delta_0$ (a result of Costas Dimitracopoulos [8], after a suggestion of Jeff Paris).[9] For the reader interested in the details of constructing one such formula, we recommend [25]. Together with bounded induction, (i) and (ii) are sufficient

[9]We thank Costas Dimitracopoulos for the attribution of this result.

to show the basic properties of the *graph* of exponentiation. Of course, what $\mathsf{I}\Delta_0$ cannot prove is its totality: $\forall x, y\, \exists z\, \mathrm{Exp}(x, y, z)$. It is easy, but instructive (the construction will be used in Theorem 7), to see why this is so. Take a non-standard model $\mathfrak{M}$ of $\mathsf{I}\Delta_0 + \exp$ (here exp is the above statement postulating the totality of the exponential function), fix a non-standard element c of the model, and consider the structure given by the *cut* $\{x \in |\mathfrak{M}| \colon \exists n \in \mathbb{N}\, (x \leq c^n)\}$ (this is a semantic cut, as opposed to the syntactic cuts that we have been using). It is clear that this cut is a model of $\mathsf{I}\Delta_0$ and that exp fails there because c^c does not exist in the cut.

The existence of the Bennett/Dimitracopoulos/Paris formula permits the definition in $\mathsf{I}\Delta_0$ of the function $x \rightsquigarrow |x| = \lceil \log(x+1) \rceil$ (this function, the binary length function, is easily seen to be provably total in $\mathsf{I}\Delta_0$). Observe that $2^{|x|}$ provably exists in $\mathsf{I}\Delta_0$ and that $\mathsf{I}\Delta_0 \vdash x > 0 \rightarrow 2^{|x|-1} < x + 1 \leq 2^{|x|}$ (we are taking, as will be usual, some liberties of expression). The axiom Ω_1 can now be introduced:

$$\Omega_1 \colon \forall x\, \exists w\, (x^{|x|} = w).$$

The formula $L(x) := \exists y\, (|y| = x)$ is called the *logarithmic* cut. It is indeed a cut and it is closed under addition in $\mathsf{I}\Delta_0$. However, we need Ω_1 to guarantee that the logarithmic cut is closed under multiplication (note that $|x| \cdot |y| = |2^{|x|\cdot|y|}|$ and that $2^{|x|\cdot|y|}$ exists in $\mathsf{I}\Delta_0 + \Omega_1$). In general, let $\omega_1(x) := x^{|x|}$ and $\omega_{n+1}(x) := 2^{\omega_n(|x|)}$, and consider the axioms $\Omega_n \colon \forall x\, \exists w\, (\omega_n(x) = w)$. It is clear that the logarithmic cut of a model of $\mathsf{I}\Delta_0 + \Omega_{n+1}$ is a model of $\mathsf{I}\Delta_0 + \Omega_n$.

Proposition 3. *$\mathsf{I}\Delta_0 + \Omega_n$ is interpretable in $\mathsf{I}\Delta_0$.*

Proof. Consider the formula $C_1(x) :\equiv \forall y\, \exists w\, (2^{|y||x|} = w)$. It is easy to show that C_1 is a cut. In order to show that C_1 is a model of $\mathsf{I}\Delta_0 + \Omega_1$ it is sufficient to argue that C_1 is closed under the function ω_1. Fix x in C_1, in order to prove that $\exists z^{C_1}\, (x^{|x|} = z)$. By hypothesis, $\forall y\, \exists w\, (2^{|y||x|} = w)$. Given an arbitrary y, $y^{|x|} < 2^{|y|\cdot|x|}$ and, therefore, $y^{|x|}$ exists. In particular, $x^{|x|}$ exists. It just remains to prove that $x^{|x|}$ lies in C_1. Fix y and let us argue that $2^{|y|\cdot|x^{|x|}|}$ exists. Since

$$|y| \cdot |x^{|x|}| < |y| \cdot |2^{|x|\cdot|x|}| = |y| \cdot |x|^2 = (|y| \cdot |x|)|x| \leq |(2|y|+1)^{|x|}| \cdot |x|$$

and given that $2^{|(2|y|+1)^{|x|}|\cdot|x|}$ exists (because x is in C_1), we get what we want.

We have shown that $\mathsf{I}\Delta_0 + \Omega_1$ is interpretable in $\mathsf{I}\Delta_0$. To generalize to Ω_n, we define the cuts by recursion. If $C_n(x)$ is a formula which defines a cut closed under ω_n, let us define the formula

$$C_{n+1}(x) :\equiv \forall y\, \big(C_n(y) \rightarrow \exists w\, \big(2_n^{|y|^{(n)}|x|^{(n)}} = w \wedge C_n(w)\big)\big),$$

where 2_n^x is the n-times iterated (base 2) exponential, i.e., $2_0^x = x$, $2_{n+1}^x = 2^{2_n^x}$, and $|y|^{(n)}$ is the n-times iteration of the length function. It can be shown that C_{n+1} is a cut in $\mathsf{I}\Delta_0$ closed under ω_{n+1}. $\dashv$

To ease syntax, Buss added new function symbols to the language of Q^+, namely $|x| := \lceil \log(x+1) \rceil$, $x \sharp y := 2^{|x||y|}$ and $\lfloor \frac{x}{2} \rfloor$. We call this extended language *Buss's language*.

DEFINITION 7. *A* sharply bounded *formula is a formula with no quantifications or just quantifications of the form* $\exists x \leq |t|$ *and* $\forall x \leq |t|$, *where* t *is a term in Buss's language where* x *does not occur.*

Observe that in the standard model of the natural numbers, an existential sharply bounded quantification can be decided by a polytime search (since the number of elements less than the length of $t(x)$, for t a term of the language of S^1_2, is polynomial in the length of x).

DEFINITION 8. *The class of* Σ^b_1*-formulas is the smallest class of formulas containing the atomic formulas and its negations, and closed under conjunctions, disjunctions, sharply bounded quantifications and bounded existential quantifications. The* Π^b_1*-formulas are defined dually, with the closure under existential bounded quantifications replaced by the closure under universal bounded quantifications.*

An existential bounded quantification corresponds to a search in exponential time because the number of elements less than x is exponential in $|x|$. This is the basic observation that explains the well-known fact that the Σ^b_1-formulas define exactly the NP-sets in the standard model of arithmetic.[10]

Buss's theory S^1_2 (which is an extension of Q^+) has the so-called BASIC axioms (consisting of 32 universal axioms, see [5] for the list) and the following form of induction: the Σ^b_1-PIND scheme $\varphi(0) \wedge \forall x\, (\varphi(\lfloor \frac{x}{2} \rfloor) \rightarrow \varphi(x)) \rightarrow \forall x\, \varphi(x)$, where φ is a Σ^b_1-formula. The choice of the above formulation of induction, instead of the more usual one (with the '+1' transition) is explained by the aim of defining a theory whose provably total functions are (in an appropriate sense) exactly the polytime computable functions. More precisely, Buss proved that (i) it is possible to extend by definitions the theory S^1_2 in order to have a function symbol for each (appropriate) description of a polytime computable function (a process that mimics the introduction of primitive recursive functions in the theory $\mathsf{I}\Sigma_1$); and (ii) if $\mathsf{S}^1_2 \vdash \forall x\, \exists y\, \varphi(x, y)$, with φ a Σ^b_1-formula, then there is a function symbol f (for a polytime computable function) such that $\forall x\, \varphi(x, f(x))$ is provable in the above referred (conservative) extension of S^1_2. Part (i) relies on some bootstrapping (with nitty-gritty details as mentioned in footnote 10), and part (ii) on an easy application of partial cut-elimination.

[10] There are some nitty-gritty details in proving this because the class of sets defined by the sharply bounded formulas is not natural from a computational complexity point of view. This is due to the specific design of the language. A language (and a theory) directly speaking of binary words (instead of natural numbers) is more natural on this regard. This was proposed in [17] (after [16]), but not generally adopted. Zambella in [50] proposes a related framework.

The interested reader can find details concerning the development of the theory S_2^1 and its main properties in the original [5] or in the textbooks [25] or [26].

PROPOSITION 4. *S_2^1 is locally interpretable in Q.*

PROOF. This follows from the fact that the theory S_2^1 is interpretable in an extension by definitions of the theory $\mathsf{I}\Delta_0 + \Omega_1$. Note that the PIND-scheme of induction follows immediately from the scheme of complete induction for bounded formulas, and this latter scheme is provable in $\mathsf{I}\Delta_0 + \Omega_1$. ⊣

The following scheme of induction is provable in S_2^1:

$$\varphi(0) \wedge \forall x\,(\varphi(|x|) \rightarrow \varphi(|x|+1)) \rightarrow \forall x\,\varphi(|x|), \tag{$*$}$$

for Σ_1^b-formulas φ. In fact, this form of induction reduces to the PIND-scheme by considering the formula $\phi(x) :\equiv \varphi(|x|)$ and using the equality $\left|\lfloor \frac{x}{2} \rfloor\right| + 1 = |x|$ $(x \neq 0)$.

PROPOSITION 5. *The theory $\mathsf{I}\Delta_0$ is interpretable in S_2^1.*

PROOF. It can be shown that the logarithmic part L is also a cut in S_2^1, closed under addition and multiplication. Moreover, a bounded quantifier, with bound in L, is a sharply bounded quantifier in the following sense: given a bounded formula $\varphi(\vec{x})$ in the language of $\mathsf{I}\Delta_0$ (free variables $\vec{x}$ as shown), the formula $\varphi(|\vec{x}|)$ is equivalent (in S_2^1) to a sharply bounded formula (the notation is self-explanatory). This result can be easily proved by induction on the complexity of the built up of the formula φ. It follows from $(*)$ that L is a model of bounded induction and, therefore, a model of $\mathsf{I}\Delta_0$. ⊣

§5. A universal predicate.

It is well-known that there is a universal recursively enumerable predicate and that, furthermore, this property is provable in a suitable subtheory of Peano Arithmetic. A similar result holds for NP-sets provided that we take into account the degrees of the polynomials of the polytime computations. In this section, we briefly review this universality result (for details, we refer the reader to pages 330–336 and 350–352 of [25]). The universality result will be used thrice in the paper. It will be used in this section to prove that the theory S_2^1 is finitely axiomatizable. It will also be used in Section 6 to show that the so-called scheme of bounded collection is interpretable in $\mathsf{I}\Delta_0 + \Omega_1$, and in Section 7 in order to code *sets*. The latter two applications only require that the universal predicate works within the theory S_2^1, but it is crucial for the first application that the universality is provable using only a *finite* fragment of S_2^1.

As mentioned in the previous section, polytime computable functions (and, hence, polytime decidable predicates) can be introduced smoothly in the theory S_2^1. There is a natural way to define a polynomial time decidable quaternary predicate $C(u, x, y, z)$ in the language of S_2^1 saying that z is the code of a terminating computation of the (deterministic) Turing machine with code u which, on input x, outputs y. This predicate is simultaneously given by a Σ_1^b and a Π_1^b formula, with their equivalence provable in S_2^1 (a predicate like this is called a Δ_1^b-predicate *in* S_2^1). Let $D(u, x, z)$ be $C(u, x, 0, z)$. It can be shown that, for any Σ_1^b-formula $\varphi(x)$, there are standard numbers $\overline{e}$ and $\overline{n}$ such that

$$\mathsf{S}_2^1 \vdash \forall x \left(\varphi(x) \leftrightarrow \exists w, z \leq 2^{(|x|+2)^{\overline{n}}} D(\overline{e}, \langle x, w \rangle, z)\right).$$

In the above, w corresponds to a non-deterministic "bounded guess." It is simple to arrange things so that e and n are the same number. With $U(u, x, l)$ defined as the formula $\exists w, z \leq l\, D(u, \langle x, w \rangle, z)$, the following can be proved:

PROPOSITION 6. *There exists a Σ_1^b-formula U which is universal in the following sense: For every Σ_1^b-formula $\phi(x)$ there exists a (standard) number e such that*

$$\mathsf{S}_2^1 \vdash \phi(x) \leftrightarrow U(\overline{e}, x, 2^{(|x|+2)^{\overline{e}}}),$$

and

$$\mathsf{S}_2^1 \vdash \forall z\, (U(\overline{e}, x, z) \rightarrow \phi(x)).$$

Furthermore, the above sentences are provable in a fixed finite fragment of S_2^1.

Some comments are in order regarding the claim that only a finite fragment of S_2^1 is needed. This is the case because Σ_1^b-formulas are obtained using only finitely many constructs (the constants and the function symbols of the language of S_2^1, equality, inequality, $\leq$, $\not\leq$, conjunction, disjunction, the two sharply bounded quantifiers and the existential bounded quantifier), and the fact that each of the corresponding constructions can be dealt appropriately within a finite fragment of S_2^1 (as we will see, this appropriateness includes the possibility of making sense of polynomials of non-standard degree). We illustrate this with the universal sharply bounded quantifier.

We have been using the expression '$2^{|x|^{\overline{n}}}$', for numerals $\overline{n}$. Of course, this makes sense because the said expression abbreviates a term of the language of S_2^1. If, instead of a numeral $\overline{n}$, we have a variable u (in model-theoretic terms, a non-standard number), then the expression does not make sense in S_2^1 any longer. Let $\tilde{C}(u, x, y, 2^{(|x|+2)^u})$ abbreviate the formula $\exists z\, (z \leq 2^{(|x|+2)^u} \wedge C(u, x, y, z))$, and $\tilde{U}(u, x, 2^{(|x|+2)^u})$ abbreviate $\exists v\, (v = 2^{(|x|+2)^u} \wedge U(u, x, v))$. Note that both expressions '$z \leq 2^{(|x|+2)^u}$' and '$v = 2^{(|x|+2)^u}$' are meaningful in S_2^1 because the *graph* of the exponentiation is polytime decidable and, in fact, given by a Δ_1^b-predicate in S_2^1. We claim that it is possible to introduce in S_2^1 a binary polytime computable function s such that the universal closure of the (single) formula

$$\tilde{U}\big(s(u,v), x, 2^{(|x|+2)^{s(u,v)}}\big) \leftrightarrow \\ \forall w, y\,\big(\tilde{C}(u,x,y,2^{(|x|+2)^u}) \wedge w \leq |y| \rightarrow \tilde{U}\big(v, \langle x, w\rangle, 2^{(|\langle x,w\rangle|+2)^v}\big)\big)$$

is provable in S^1_2. A complete proof of this claim would take a great deal of work but the idea is clear. It hinges on the fact that from (codes of) Turing machines u and v one can produce a Turing machine (code) $s(u,v)$ that runs as follows: on input x, it first computes (within appropriate time) y according to the Turing machine (code) u and then, for each $w \leq |y|$, it verifies that the Turing machine (code) v accepts (after a suitable bounded guess and in appropriate time) the pair $\langle x, w\rangle$. The production of the code $s(u,v)$ and the verification that it does what was just described can be formalized in S^1_2.

For each of the finitely many constructs needed to obtain Σ^b_1-formulas, a correspondingly single property is provable in S^1_2. Only a finite fragment of S^1_2 is needed to obtain these properties. They are then used to obtain the formal proofs claimed to exist in Proposition 6.

COROLLARY 1. *Buss's theory* S^1_2 *is finitely axiomatizable.*

PROOF. Let F consist of all the 32 BASIC axioms of S^1_2, finitely many PIND-induction axioms needed to prove the universality conditions of the previous proposition (we also assume that F is strong enough to have a pairing operation and prove some simple inequalities), together with the universal closure of the following (single) formula:

$$U(u, \langle 0, y\rangle, l) \wedge \neg U(u, \langle x, y\rangle, l) \\ \rightarrow \exists w < x\,\big(U(u, \langle \lfloor \tfrac{w}{2} \rfloor, y\rangle, l) \wedge \neg U(u, \langle w, y\rangle, l)\big).$$

Note that the above sentence is provable in S^1_2 (by a suitable PIND-induction on x with parameters u, y and l). It is not difficult to see, using both conditions of Proposition 6, that every form of Σ^b_1-induction can be deduced in F. $\dashv$

By Proposition 4, we know that S^1_2 is locally interpretable in Q. Since we have just proved that S^1_2 is finitely axiomatizable, we conclude that S^1_2 is (globally) interpretable in Q. The next result follows easily:

THEOREM 3. $\mathsf{I}\Delta_0 + \Omega_n$ *is interpretable in* Q.

PROOF. By Proposition 3, it is enough to show that $\mathsf{I}\Delta_0$ is interpretable in Q. This follows from the above discussion and the fact that $\mathsf{I}\Delta_0$ is interpretable in S^1_2 (see Proposition 5). $\dashv$

§6. **The interpretability of bounded collection in** Q**.** We denote by $\mathsf{B}\Sigma_1$ the scheme of *bounded collection*:

$$\forall x \leq z\, \exists y\, \varphi(x, y) \rightarrow \exists w\, \forall x \leq z\, \exists y \leq w\, \varphi(x, y),$$

where φ is a bounded formula, possibly with parameters. The following result, in the local interpretability version, can be traced to chapter 22 of [28] (see [43] and [24] for alternative proofs of the global version):

PROPOSITION 7. $\mathsf{I}\Delta_0 + \mathsf{B}\Sigma_1$ *is cut-interpretable in* $\mathsf{I}\Delta_0 + \Omega_1$.

PROOF. By Proposition 6, we know that if ψ is a sharply bounded formula, then

$$\mathsf{I}\Delta_0 + \Omega_1 \vdash \forall x, y\, (\psi(x, y, p) \leftrightarrow \exists z\, U(e, \langle x, y, p\rangle, z))$$

where we are using a standard coding of triples $w = \langle w_0, w_1, w_2\rangle$. Let $J(u)$ be the formula defined as the conjunction of $L(u)$ (i.e., u is in the logarithmic cut) together with:

$$\forall e, p\, [\forall x \leq u\, \exists y\, U(e, \langle x, y_0, p\rangle, y_1) \\ \rightarrow \exists w\, \forall x \leq u\, \exists y \leq w\, U(e, \langle x, y_0, p\rangle, y_1)].$$

It is clear that, given any sharply bounded formula $\psi(x, y, p)$ and u in the cut J, we have:

$$\forall x \leq u\, \exists y\, \psi(x, y, p) \rightarrow \exists w\, \forall x \leq u\, \exists y \leq w\, \psi(x, y, p). \qquad (**)$$

The above also holds for Σ_1^b-formulas ψ (just use pairing). We claim the following:

1. $J(u) \wedge u' \leq u \rightarrow J(u')$,
2. $J(u) \rightarrow J(\mathsf{S}u)$.

For the first property, apply $(**)$ to the formula

$$x \leq u' \rightarrow U(e, \langle x, y_0, p\rangle, y_1).$$

The fact that J is inductive is obvious. Now, by shortening, take I a subcut of J closed under addition and multiplication. As we know, I is a model of $\mathsf{I}\Delta_0$. It remains to see that $\mathsf{B}\Sigma_1$ also holds in I. Consider $\varphi(x, y)$ a bounded formula in the language of $\mathsf{I}\Delta_0$ (possibly with parameters in I), take u in I, and assume that $\forall x \leq u\, \exists y^I\, \varphi(x, y)$. The goal is to prove that there is w in I such that $\forall x \leq u\, \exists y \leq w\, \varphi(x, y)$. We use the fact that a bounded formula of the language of $\mathsf{I}\Delta_0$, with parameters in the logarithmic part of a model of $\mathsf{I}\Delta_0 + \Omega_1$, is equivalent to a sharply bounded formula (see the proof of Proposition 5). It is now clear, by property $(**)$, that there is w such that $\forall x \leq u\, \exists y \leq w\, \varphi(x, y)$. Let w_0 be the least such w. We prove that w_0 is in I. Obviously, by minimality, there is $x_0 \leq u$ such that $\forall y < w_0\, \neg\varphi(x_0, y)$. By hypothesis, take y_0 in I such that $\varphi(x_0, y_0)$. Thus $w_0 \leq y_0$. Since y_0 is in I and I is a cut, we conclude that w_0 lies in I. ⊣

THEOREM 4. $\mathsf{I}\Delta_0 + \Omega_n + \mathsf{B}\Sigma_1$ *is interpretable in* Q.

PROOF. By the above proposition, $\mathsf{I}\Delta_0 + \mathsf{B}\Sigma_1$ is interpretable in $\mathsf{I}\Delta_0 + \Omega_1$ on a cut. Just shorten this cut to a cut closed under ω_n. Note that $\mathsf{I}\Delta_0 + \Omega_1$ is interpretable in Q. ⊣

§7. The interpretability of a portion of analysis in Q**.** The theory BTPSA (an acronym for "Base Theory for Polynomial Space Analysis") was introduced in [20], where it was proved that its provably total functions (with appropriate graphs) are exactly the functions computable in polynomial space. As mentioned in the introduction, in BTPSA it is possible to develop a *modicum* of analysis (e.g., it is shown in [21, 19] that Riemann integration for functions with a modulus of uniform continuity can be developed in BTPSA up to the fundamental theorem of calculus). In this section, we describe the system BTPSA in unary notation (instead of the original binary notation adopted in [20]) and show how it can be interpreted in Q.

Let $\mathcal{L}_2$ be the second-order language whose first-order part is Buss's language (i.e., the language of the theory S^1_2). The second-order part consists of set variables $X, Y, Z, \dots$ and attending quantifiers. There is also a binary relation symbol '$\in$' that infixes between first-order terms and set variables. We may think of $\mathcal{L}_2$ as a two-sorted first-order language: one sort for numbers, the other for sets of numbers.

DEFINITION 9. *A* $\Sigma^{1,b}_0$*-formula is a formula of* $\mathcal{L}_2$ *without second-order quantifications and where all the first-order quantifications are bounded. It may have second-order parameters.*

In the sequel, we need to speak of second-order bounded quantifiers. These are written like $\forall X \leq t\,(\dots)$ or $\exists X \leq t\,(\dots)$, where X is a second-order variable and t is a (first-order) term. The idea is that we are restricting second-order quantifications to subsets of the initial segment determined by the term t. This restriction can be written down explicitly but, for reasons of complexity, it is better to formulate it in an alternative way. We consider the bounded quantification $\forall X \leq t\,\varphi(X)$ as an abbreviation of the formula $\forall X\,\varphi(X^t)$, where X^t is a way of indicating that, whenever a complex of the form '$q \in X$' occurs in $\varphi(X)$, then it should be replaced by '$q \in X \wedge q \leq t$'. *Mutatis mutandis* for second-order bounded existential quantifications.

DEFINITION 10. *A* $\Sigma^{1,b}_\infty$*-formula is a formula of* $\mathcal{L}_2$ *where all the first and second-order quantifications are bounded. It may have second-order parameters.*

The following two sub-classes of $\Sigma^{1,b}_\infty$-formulas are important:

DEFINITION 11. *A* $\Sigma^{1,b}_1$*-formula is a formula of* $\mathcal{L}_2$ *of the form* $\exists X \leq t\,\varphi$*, where* φ *is a* $\Sigma^{1,b}_0$*-formula. The* $\Pi^{1,b}_1$*-formulas are defined dually.*

It is easy to argue that we have contraction for blocks of quantifiers of the form $\exists X \leq t$ (and, dually, for blocks of quantifiers $\forall X \leq t$). We can now define the theory BTPSA. The first three classes of axioms are clear enough. The last two classes are forms of bounded collection.

DEFINITION 12. BTPSA *is the theory in* $\mathcal{L}_2$ *with the following axioms*:

- *the* BASIC *axioms*;
- *the* PIND (*induction*) *axioms for extended* $\Sigma_1^{1,b}$*-formulas*;[11]
- *the following form of comprehension*:

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

  where φ *is a* $\Sigma_1^{1,b}$*-formula and* ψ *is a* $\Pi_1^{1,b}$*-formula*;
- $B\Sigma_\infty^{1,b}$*-collection*:

$$\forall x \leq t\,\exists y\,\varphi(x,y) \rightarrow \exists z\,\forall x \leq t\,\exists y \leq z\,\varphi(x,y),$$

  where φ *is a* $\Sigma_\infty^{1,b}$*-formula*;
- $B^1\Sigma_\infty^{1,b}$*-collection*:

$$\forall X \leq t\,\exists y\,\varphi(y,X) \rightarrow \exists z\,\forall X \leq t\,\exists y \leq z\,\varphi(y,X),$$

  where φ *is a* $\Sigma_\infty^{1,b}$*-formula.*

The reader familiar with bounded second-order systems should notice that the above PIND-axioms correspond exactly to the induction present in Buss's system U_2^1 (a system related to polynomial space computability). The comprehension scheme may be a bit surprising on a first reading since the antecedent is strong enough to obtain all the recursive sets in the standard model (way beyond the polynomial space decidable sets). The system BTPSA is, nevertheless, Π_2^0-conservative over Buss's U_2^1 (see [20]).[12]

THEOREM 5. BTPSA *is interpretable in* Q.

PROOF. We show that BTPSA is interpretable in $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$. We assume that the language of the latter theory is Buss's language. Note that

$$\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1 \vdash \forall w_1,\dots,w_n\,\exists u\,(|u| = t(|w_1|,\dots,|w_n|))$$

holds for every term t of the language. This is because our theory has the axiom Ω_2. In the sequel, this fact will be used many times without explicit mentioning.

Let us now describe the interpretation. The number sort of BTPSA is interpreted by the logarithmic cut L (which is a model of $\mathsf{I}\Delta_0 + \Omega_1$). The set

[11] The class of *extended* $\Sigma_1^{1,b}$*-formulas* is the smallest class of formulas containing the $\Sigma_0^{1,b}$-formulas, and closed under conjunctions, disjunctions, bounded first-order quantifications and existential second-order bounded quantifications.

[12] A similar phenomenon occurs with the well-known system RCA_0 of reverse mathematics. This system has recursive comprehension but it is Π_2^0-conservative over primitive recursive arithmetic (see [38]). Intuitively, in order to form a set using a comprehension scheme of the above form, one has to have the antecedent available. The weaker the theory, the lesser the availability.

sort is interpreted by elements α that satisfy the condition $\mathrm{Set}(\alpha)$ given by:

$$\forall u, x\,\big(|u| = x \to \big[\exists w\; U(\alpha_0, \langle u, w_0, \alpha_1\rangle, w_1) \\ \leftrightarrow \forall w\,\neg U(\alpha_2, \langle u, w_0, \alpha_3\rangle, w_1)\big]\big),$$

where α codes the quadruple $\langle \alpha_0, \alpha_1, \alpha_2, \alpha_3\rangle$ (w is seen as the pair $\langle w_0, w_1\rangle$). Here, U is the universal Σ^b_1-predicate of Section 5. Note that a single element may play a double role: one as a number, the other as a set (overlapping pieces, as mentioned in the introduction). The first-order vocabulary is, by definition, unaltered by the interpretation (on the first-order part, we are describing a cut-interpretation). The membership operation is defined thus:

$$x \in \alpha :\equiv \exists u, v, z\,(|u| = x \wedge U(\alpha_0, \langle u, v, \alpha_1\rangle, z)).$$

It is given by an existential claim applied to a Σ^b_1-formula: a so-called $\exists\Sigma^b_1$-formula. If $\mathrm{Set}(\alpha)$ then, for x in the logarithmic part, the above condition is equivalent to $\forall u, v, z\,(|u| = x \to \neg U(\alpha_2, \langle u, v, \alpha_3\rangle, z))$. Therefore, in this case, '$x \in \alpha$' is also given by a (so-called) $\forall\Pi^b_1$-formula.

The above specifications define a translation $\varphi \rightsquigarrow \varphi^*$ from formulas of the second-order language of BTPSA to formulas of Buss's first-order language. For instance, $(\forall x\,\varphi(x))^*$ is $\forall x^L\,\varphi^*(x)$. The translation of a set-quantification needs no comments except that we write $(\forall X\,\varphi(X))^*$ as $\forall\alpha\,(\mathrm{Set}(\alpha) \to \varphi^*(\alpha))$: capital Latin letters are changed to Greek lower case letters. For ease of reading, we even write $\forall\alpha\,\varphi^*(\alpha)$, but the antecedent $\mathrm{Set}(\alpha)$ is supposed to be always implicit. Finally, the translation of $\forall X \le t\,\varphi(X)$ is written as the translation of $\forall X\,\varphi(X^t)$. We use corresponding notations for existential quantifications.

In order to show that the translations of the axioms of BTPSA are provable in $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$, we rely on a few facts.

Fact 1. *If φ and ψ are bounded formulas, then the theory $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$ proves*

$$\forall x \le z\,(\exists y\,\varphi(x, y) \leftrightarrow \forall y\,\psi(x, y)) \\ \to \exists w\,\forall x \le z\,(\exists y\,\varphi(x, y) \leftrightarrow \exists y \le w\,\varphi(x, y)).$$

Proof of Fact 1. The antecedent of the above formula implies

$$\forall x \le z\,\exists u, y\,(\psi(x, u) \to \varphi(x, y)).$$

By $\mathsf{B}\Sigma_1$, there is w such that $\forall x \le z\,\exists u, y \le w\,(\psi(x, u) \to \varphi(x, y))$. It is easy to see that this w does the job. ⊣

Fact 2. *The theory $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$ proves*

$$\forall\alpha\,\forall z\,\exists y < 2^{|z|+1}\,\forall x \le |z|\,(y_x = 1 \leftrightarrow x \in \alpha),$$

where y_x denotes the xth-bit of the binary expansion of y.

PROOF OF FACT 2. Let α be such that $\mathrm{Set}(\alpha)$ and fix z. As we know, $x \in \alpha$ is, simultaneously, of the form $\exists\Sigma_1^b$ and $\forall\Pi_1^b$. Therefore, so is the equivalence $y_x = 1 \leftrightarrow x \in \alpha$. By the previous fact, this equivalence (for $x \le |z|$) can be considered a bounded formula. It is now clear that we can argue by induction to obtain the above fact. $\dashv$

FACT 3. *Given $\varphi(x_1, \dots, x_n)$ a $\Sigma_0^{1,b}$-formula (with the first-order free variables as shown), there is a $\exists\Sigma_1^b$-formula $\phi(w_1, \dots, w_n)$ such that*

$$\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1 \vdash \forall w_1 \dots \forall w_n\, (\varphi^*(|w_1|, \dots, |w_n|) \leftrightarrow \phi(w_1, \dots, w_n)).$$

There is also a $\forall\Pi_1^b$-formula with the same property.

PROOF OF FACT 3. Intuitively, if we consider $x_1, \dots, x_n$ as first-order parameters then, when the interpretation acts, they become logarithmic elements, i.e., elements of the form $|w_1|, \dots, |w_n|$ and, consequently, a first-order bounded quantification becomes a sharply bounded quantification. Let us turn to a precise argument. Up to logical equivalence, $\Sigma_0^{1,b}$-formulas can be obtained from atomic formulas and their negations ('$t \in X$' and '$t \notin X$' are included), using conjunction, disjunction and first-order bounded quantifications. Let us consider the case when $\varphi(x_1, \dots, x_n)$ is $t(x_1, \dots, x_n) \in X$. Then, $\varphi^*(|w_1|, \dots, |w_n|)$ is

$$\exists u, v, z\, (|u| = t(|w_1|, \dots, |w_n|) \wedge U(\alpha_0, \langle u, v, \alpha_1 \rangle, z)).$$

The above formula is $\exists\Sigma_1^b$. The other atomic cases are also easy and it is clear that the above fact is preserved by conjunctions and disjunctions. Let us now consider a universal (first-order) bounded quantification

$$\forall x \le t(x_1, \dots, x_n)\, \varphi(x, x_1, \dots, x_n).$$

By induction hypothesis, there is a Σ_1^b-formula $\phi'(z, w, w_1, \dots, w_n)$ such that the following equivalence holds in $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$:

$$\varphi^*(|w|, |w_1|, \dots, |w_n|) \leftrightarrow \exists z\, \phi'(z, w, w_1, \dots, w_n).$$

Now, the formula $\forall x \le t(|w_1|, \dots, |w_n|)\, \varphi^*(x, |w_1|, \dots, |w_n|)$ is equivalent to

$$\exists u\, (|u| = t(|w_1|, \dots, |w_n|) \\ \wedge \forall x \le |u|\, \exists z, w\, (|w| = x \wedge \phi'(z, w, w_1, \dots, w_n))).$$

By bounded collection $\mathsf{B}\Sigma_1$, we get

$$\exists y, u\, (|u| = t(|w_1|, \dots, |w_n|) \\ \wedge \forall x \le |u|\, \exists z, w \le y\, (|w| = x \wedge \phi'(z, w, w_1, \dots, w_n))).$$

This is a $\exists\Sigma_1^b$-formula. The case of existential (first-order) bounded quantification is even simpler. A dual argument establishes the remaining cases. $\dashv$

FACT 4. *Given $\varphi(x_1, \dots, x_n)$ a $\Sigma_1^{1,b}$-formula (with the first-order free variables as shown), there is a $\exists\Sigma_1^b$-formula $\psi(w_1, \dots, w_n)$ such that*

$$\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1 \vdash \forall w_1 \dots \forall w_n\, (\varphi^*(|w_1|, \dots, |w_n|) \leftrightarrow \psi(w_1, \dots, w_n)).$$

The dual property (i.e., the existence of an appropriate $\forall\Pi_1^b$-formula) holds for $\Pi_1^{1,b}$-formulas.

PROOF OF FACT 4. A $\Sigma_1^{1,b}$-formula is of the form $\exists X \leq t(x)\, \varphi$ with φ a $\Sigma_0^{1,b}$-formula. For readability, we took the bounding term with only a single variable x and we omitted the parameters of φ. By Fact 2, $\exists\alpha\, \varphi^*(\alpha^{t(|w|)})$ is equivalent to

$$\exists u\, \big(|u| = t(|w|) \wedge \exists y < 2^{|u|+1}\, \varphi^*(y^{|u|})\big),$$

where $\varphi^*(y^{|u|})$ is obtained from $\varphi^*(\alpha^{|u|})$ by substituting complexes of the form '$r \in \alpha \wedge r \leq |u|$' by the complexes '$y_r = 1 \wedge r \leq |u|$'. By Fact 3, the formula $\varphi^*(\alpha^{|u|})$ is equivalent to a $\exists\Sigma_1^b$-formula. It is not difficult to see that the same holds for the formula $\varphi^*(y^{|u|})$. The result follows. The dual property is similar. $\dashv$

FACT 5. *Given $\varphi(x_1, \dots, x_n)$ a $\Sigma_\infty^{1,b}$-formula (with the first-order free variables as shown), there is a bounded formula $\phi(y, w_1, \dots, w_n)$ such that*

$$\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1 \vdash \forall w_1 \dots \forall w_n\, (\varphi^*(|w_1|, \dots, |w_n|) \leftrightarrow \exists y\, \phi(y, w_1, \dots, w_n)).$$

A similar property holds with a universal quantifier '$\forall y$' instead of '$\exists y$'.

PROOF OF FACT 5. The proof is done by induction on the complexity of the formula, following the blueprint of the above proof (second-order bounded quantifications are translated by first-order bounded quantifications). Fact 2 and $\mathsf{B}\Sigma_1$ are used systematically. $\dashv$

We are ready to see that the translations of the axioms of BTPSA are provable in the theory $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$. This is immediate for the BASIC axioms. Now, consider an extended $\Sigma_1^{1,b}$-formula φ and an element z_0 in L such that $\varphi(0) \wedge \neg\varphi(z_0)$. By Fact 5, the formula $\varphi^*(|w|)$ is equivalent to both a formula of the form $\exists y\, \phi_1(y, w)$ and a formula of the form $\forall y\, \phi_2(y, w)$, with ϕ_1 and ϕ_2 bounded formulas. Hence, by Fact 1, $\varphi^*(|w|)$ is equivalent to a bounded formula as long as $w \leq w_0$, for w_0 such that $|w_0| = z_0$. Since $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$ includes bounded induction, it follows that the translations of the $\Sigma_1^{1,b}$-PIND axioms hold in $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$.

Let us study the translations of the comprehension axioms of BTPSA. Assume $\forall x^L\, (\exists y^L\, \varphi^*(x, y, p_0) \leftrightarrow \forall y^L\, \psi^*(x, y, p_1))$, where φ is a $\Sigma_1^{1,b}$-formula, ψ is a $\Pi_1^{1,b}$-formula and where we show explicitly the parameters p_0 and p_1 (without loss of generality, they are single parameters; for simplicity, we omit second-order parameters). Note that p_0 and p_1 lie in the logarithmic part (say, with $|q_0| = p_0$ and $|q_1| = p_1$). By Fact 4, it follows that $\exists y^L\, \varphi^*(x, y, p_0)$ is equivalent to $\exists v\, \varphi'(u, v, q_0)$, where $|u| = x$ and φ'

is a Σ_1^b-formula. Similarly, $\forall y^L\, \psi^*(x, y, p_1)$ is equivalent to $\forall v\, \psi'(u, v, q_1)$, this time with ψ' a Π_1^b-formula. By Proposition 6, there are (standard) numbers d and e such that S_2^1 (and, hence, $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$) proves the equivalences

$$\varphi'(u, v, q_0) \leftrightarrow U\big(\overline{d}, \langle u, v, q_0\rangle, 2^{|\langle u,v,q_0\rangle + 2|^{\overline{d}}}\big) \text{ and}$$
$$\psi'(u, v, q_1) \leftrightarrow \neg U\big(\overline{e}, \langle u, v, q_1\rangle, 2^{|\langle u,v,q_1\rangle + 2|^{\overline{e}}}\big).$$

It is now clear that we can take the quadruple $\alpha := \langle \overline{d}, q_0, \overline{e}, q_1\rangle$ to code the set required by the comprehension axiom.

Finally, using the Facts, it is easy to see that both bounded collection schemes translate into forms of bounded collection that are available in $\mathsf{I}\Delta_0 + \Omega_2 + \mathsf{B}\Sigma_1$. ⊣

The interpretation above does more than one was aiming. For instance, it interprets the usual "plus one" kind of extended $\Sigma_1^{1,b}$-induction (for the expert, we obtain a theory which is Π_2^0-conservative over Buss's theory V_2^1). With this kind of induction, it is possible to introduce every exponential time function (with polynomial growth rate). Actually, it interprets $\Sigma_\infty^{1,b}$-induction and, with a small adaptation, it interprets also comprehension for every fixed level of (dual) formulas φ and ψ (see the statement of comprehension in Definition 12) of the $\Sigma_\infty^{1,b}$-hierarchy (the hierarchy which counts the number of alternations of second-order bounded quantifications). For the reader of like mind, one can even envisage interpreting comprehension for all the second-order bounded hierarchy at once by considering an *interpreted* theory with the axiom Ω_3. Of course, more can be obtained by going up in the hierarchy of the Ω_n's.

It has been known since the nineties that forms of weak König's lemma (see [38] for a classical discussion of this principle) are Π_2^0-conservative over theories of bounded arithmetic (they are even first-order conservative if the theories of bounded arithmetic include corresponding forms of bounded collection). The first paper on this subject was [18], but there has been others like [7] or [13]. Along the years there has been investigations of weak analysis with forms of weak König's lemma, as witnessed by the Ph.D. theses of António Fernandes and the second author (see [12] and [21]). Recently, António Fernandes published in [14] a study concerning a strong form of weak König's lemma, namely a version of the so-called *strict* Π_1^1-*reflection* principle:[13]

$$\forall X\, \exists x\, \varphi(X, x) \rightarrow \exists z\, \forall X\, \exists x \leq z\, \varphi(X, x),$$

[13]This principle originally appeared in the context of admissible set theory (see [1]). It has been observed by several people (notoriously by Gerald Sacks in the preface of his [36]) that fragments of arithmetic have similarities with admissibility.

where φ is a $\Sigma^{1,b}_{\infty}$-formula. It is known that this principle is first-order conservative over BTPSA. We conjecture that this principle is interpretable in Q.

§**8. Negative results.** We saw that various, successively stronger, theories are interpretable in Q. In this section, we show that there are limits for this exercise.

LEMMA 2. *The theory* $\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q})$ *is not interpretable in* Q.

PROOF. Suppose, in order to reach a contradiction, that $\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q})$ is interpretable in Q. Given that $\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q})$ is finitely axiomatizable (Corollary 1), $\mathsf{S}^1_2 \vdash \mathrm{Con}(\mathsf{Q}) \rightarrow \mathrm{Con}(\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q}))$. Hence, $\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q}) \vdash \mathrm{Con}(\mathsf{S}^1_2 + \mathrm{Con}(\mathsf{Q}))$. This contradicts Gödel's second incompleteness theorem. ⊣

Consider the formula $\mathrm{SL}(x) :\equiv \exists y\, (2^2_x = y)$. In $\mathsf{I}\Delta_0 + \exp$, the formula SL is inductive and, hence, defines a cut: the so-called super-logarithmic cut. By Proposition 1, there is a subcut $J(x)$ of the super-logarithmic cut closed under addition and multiplication. Without loss of generality (by further shortening), we may suppose that J is a model of $\mathsf{I}\Delta_0 + \Omega_1$. The following is easy: if x is in J then 2^x_x exists. The proof of this fact lies in the observation that if 2^w_{z+1} exists then 2^{w+1}_z also exists and $2^{w+1}_z \leq 2^w_{z+1}$. Given that $2x - 2$ lies in J, 2^2_{2x-2} exists ($x \neq 0$). Then, by (bounded) induction, it is clear that 2^x_x exists and $2^x_x \leq 2^2_{2x-2}$.

The following result first appeared in [48]:

LEMMA 3 (Wilkie). $\mathsf{I}\Delta_0 + \Omega_1 + \mathrm{Con}(\mathsf{Q})$ *is interpretable in* $\mathsf{I}\Delta_0 + \exp$.

PROOF. (Sketch) For technical reasons, we prove the equivalent claim that $\mathsf{I}\Delta_0 + \Omega_1 + \mathsf{Con}(\mathsf{Q}^P)$ is interpretable in $\mathsf{I}\Delta_0 + \exp$ (Q^P is the open theory defined in Section 3). Let J be a subcut of the super-logarithmic cut which is a model of $\mathsf{I}\Delta_0 + \Omega_1$. We claim that there is no Hilbert-style proof of $0 = 1$ from the axioms of Q^P in the cut J. Let us see why. Suppose that there is such a proof in J. Then, in this very cut, there is a pure logic Hilbert-style proof of the conditional '$\forall \vec{x}\, A(\vec{x}) \rightarrow 0 = 1$' where $\forall \vec{x}\, A(\vec{x})$, with $A(\vec{x})$ quantifier-free, is (equivalent to) the conjunction of the axioms of Q^P (including the finitely many identity axioms). Note that this pure logic proof lies in J (the theorem of deduction is formalizable in $\mathsf{I}\Delta_0 + \Omega_1$). By known analyses of the proof of Gentzen's cut-elimination theorem, there is enough room to produce a cut-free proof of $\exists \vec{x}\, (A(\vec{x}) \rightarrow 0 = 1)$ (the cut-free proof has a super-exponential bound with respect to the length of the original proof). It is well-known that, from a cut-free proof of the existential sentence $\exists \vec{x}\, (A(\vec{x}) \rightarrow 0 = 1)$, we can extract closed terms $\vec{t_1}, \ldots, \vec{t_n}$ of the language of Q^P such that the conditional $A(\vec{t_1}) \wedge \cdots \wedge A(\vec{t_n}) \rightarrow 0 = 1$ is a tautology. (Note that the argument from the cut-free proof to the above tautology is formalizable in $\mathsf{I}\Delta_0 + \exp$.) This is impossible because the theory

$\mathsf{I}\Delta_0 + \exp$ is able to define a truth predicate for closed atomic formulas of the language of Q^P.

We have argued that J is a model of $\mathsf{I}\Delta_0 + \Omega_1 + \mathsf{Con}(\mathsf{Q}^P)$. Therefore, the theory $\mathsf{I}\Delta_0 + \Omega_1 + \mathsf{Con}(\mathsf{Q}^P)$ is interpretable in $\mathsf{I}\Delta_0 + \exp$. ⊣

THEOREM 6. $\mathsf{I}\Delta_0 + \exp$ *is* not *interpretable in* Q.

PROOF. This is an immediate consequence of the above two lemmas and the fact that S^1_2 is interpretable in $\mathsf{I}\Delta_0 + \Omega_1$. ⊣

In $\mathsf{I}\Delta_0 + \Omega_1$, the super-logarithmic part is no longer inductive. To see this, consider a model of $\mathsf{I}\Delta_0 + \Omega_1$ in which exp fails. Let w be such that 2^w does not exist. Take the maximum x such that $2^2_x \leq w$ ($\mathsf{I}\Delta_0 + \Omega_1$ can prove the existence of this maximum). Therefore, there is a unique x (it is the previous x or its successor) such that

$$\exists y\,(y = 2^2_x \wedge \forall z, w\,(w = 2^2_z \rightarrow z \leq x)).$$

Note that this (unique) x is in the super-logarithmic part, but its successor is not. In short, in a model in which exp fails, its super-logarithmic part has a maximum.

The following result and elegant argument are due to Solovay (see [6]):

PROPOSITION 8 (Solovay). *There are sentences ϕ_0 and ϕ_1 such that both theories $\mathsf{I}\Delta_0 + \Omega_1 + \phi_0$ and $\mathsf{I}\Delta_0 + \Omega_1 + \phi_1$ are cut-interpretable in $\mathsf{I}\Delta_0 + \Omega_2$ but the theory $\mathsf{I}\Delta_0 + \Omega_1 + (\phi_0 \wedge \phi_1)$ is not interpretable in there.*

PROOF. Let ϕ_{even} be the sentence

$$\exists y, x\,(y = 2^2_x \wedge \forall z, w\,(w = 2^2_z \rightarrow z \leq x) \wedge \text{“}x \text{ is even”}),$$

(the super-logarithmic part has a maximum and it is even) and ϕ_{odd} be the sentence

$$\exists y, x\,(y = 2^2_x \wedge \forall z, w\,(w = 2^2_z \rightarrow z \leq x) \wedge \text{“}x \text{ is odd”}),$$

(the super-logarithmic part has a maximum and it is odd). Now, let us define $\phi_0 :\equiv \neg \exp \rightarrow \phi_{even}$ and $\phi_1 :\equiv \neg \exp \rightarrow \phi_{odd}$. It should be clear that the cuts $I_0(x) :\equiv (\neg \exp \wedge \neg\phi_{even}) \rightarrow L(x)$ and $I_1(x) :\equiv (\neg \exp \wedge \neg\phi_{odd}) \rightarrow L(x)$ are, respectively, models of ϕ_0 and ϕ_1 (L is the logarithmic cut). Let us check the case I_0. If exp holds then the cut I_0 is all the model and obviously ϕ_0 holds in there. If exp fails we have two cases. If ϕ_{even}, then the cut I_0 is again all the model and, by hypothesis, ϕ_0 holds in there. If $\neg\phi_{even}$, then we must have ϕ_{odd} (because exp fails) and the cut is the logarithmic part L of the model. But, of course, ϕ_{even} holds in L (and exp continues to fail in L). We conclude that ϕ_0 holds in L, as wanted.

Observe that $(\phi_0 \wedge \phi_1) \rightarrow \exp$. By the previous theorem, $\mathsf{I}\Delta_0 + \Omega_1 + (\phi_0 \wedge \phi_1)$ is not interpretable in Q. ⊣

THEOREM 7 (Solovay). $\mathsf{I}\Delta_0 + \neg\exp$ is *interpretable in* Q.

PROOF. (Sketch) We show that $\mathsf{I}\Delta_0 + \neg\exp$ is interpretable in $\mathsf{I}\Delta_0 + \Omega_1$ (and, hence, in Q) by internalizing the argument after Proposition 2. The particular internalization depends on whether the scheme of Σ_2-induction holds, or not, in the ambient model.

It easily follows from the work in Section 5 that there is a universal Σ_2-predicate $V(e, x, p)$ in the following precise sense: for every formula $\varphi(x)$ of the form $\exists y\, \forall z\, \rho(x, y, z)$, with ρ a Σ_1^b-formula (possibly with parameters), there is a standard number e such that $\mathsf{I}\Delta_0 + \Omega_1 \vdash \exists p\, \forall x\, (\varphi(x) \leftrightarrow V(e, x, p))$. Let θ be the formula:

$$\forall p, e\, \forall x\, (V(e, 0, p) \wedge \forall x\, (V(e, x, p) \to V(e, \mathrm{S}x, p)) \to \forall x\, V(e, x, p)).$$

Note that θ holds if, and only if, the scheme of Σ_2-induction holds. We show how to interpret $\mathsf{I}\Delta_0 + \neg\exp$ (into $\mathsf{I}\Delta_0 + \Omega_1$) depending on whether one θ holds or not. This case distinction does not pose a problem for obtaining the interpretation: the interpretation effectively takes the form "if θ, proceed according to A, if not proceed according to B."

Suppose that θ holds. Of course, $\mathsf{I}\Delta_0 + \Omega_1 + \theta$ is just first-order Peano Arithmetic with induction restricted to Σ_2-formulas, the so-called theory $\mathsf{I}\Sigma_2$. We show that $\mathsf{I}\Delta_0 + \neg\exp$ is interpretable in $\mathsf{I}\Sigma_2$. Firstly, observe that $\mathsf{I}\Sigma_2$ proves that the theory $\mathsf{I}\Delta_0 + \exp + \{\overline{n} < c : n \in \mathbb{N}\}$ (framed in the language of arithmetic with an extra constant c) is consistent.[14] This is because every finite part (in the sense of the ambient theory $\mathsf{I}\Sigma_2$) of $\mathsf{I}\Delta_0 + \exp + \{\overline{n} < c : n \in \mathbb{N}\}$ has an appropriate *weak model* (see section II.8 of [38], adapted to a setting using explicit definitions instead of second-order variables). The weak model can be taken as a model in the sense that (the universal closures of) bounded formulas (with a primitive symbol for the exponentiation function) are true with respect to a suitable (partial) truth predicate. Then, by a cut-elimination argument (see, again, [38]), we conclude that this finite part is consistent. Therefore, $\mathsf{I}\Delta_0 + \exp + \{\overline{n} < c : n \in \mathbb{N}\}$ is consistent. By the formalized Gödel's completeness theorem (a.k.a. the Hilbert–Bernays completeness theorem), there is a Δ_2-definable model of the theory $\mathsf{I}\Delta_0 + \exp + \{\overline{n} < c : n \in \mathbb{N}\}$ in $\mathsf{I}\Sigma_2$ (see [39] and, also, [10] or section IV.3 of [38]). Note that the Hilbert–Bernays completeness theorem is formalizable in $\mathsf{I}\Sigma_2$. Therefore, we get an interpretation of $\mathsf{I}\Delta_0 + \exp + \{\overline{n} < c : n \in \mathbb{N}\}$. By the opening paragraph of this proof, it is clear now how to get an interpretation in which exp fails.

[14]It is known that the theory $\mathsf{I}\Delta_0 + \exp$ is finitely axiomatizable (if we disregard the language, note that it is the same theory as the finitely axiomatizable $\mathsf{S}_2^1 + \exp$) and, therefore, its formalization in $\mathsf{I}\Sigma_2$ poses no problems (in any case, the usual infinite axiomatization does not pose special problems either). The part $\{\overline{n} < c : n \in \mathbb{N}\}$ is formalized as consisting of the numbers of the form $\ulcorner \dot{x} \leq c \urcorner$ (see [39] for the dot notation and the square corners).

Suppose that θ fails. Then there is e_0 and p_0 such that the following is a *proper* cut:

$$C(x) :\equiv V(e_0, 0, p_0) \\ \wedge \forall z\, (V(e_0, z, p_0) \rightarrow V(e_0, \mathrm{S}z, p_0)) \rightarrow \forall y \leq x\, V(e_0, y, p_0).$$

Take b outside of C and, by shortening, take J a subcut of C which is a model of $\mathsf{I}\Delta_0$. If exp fails in J, we have produced the required interpretation. Otherwise, let a be the length of b (i.e., $a = |b|$). Consider the cut $K(x) :\equiv \exists u^J\, (x \leq a^u)$. Note that for u in J, a^u is defined. This is because $|a|$ does not lie in J (this stems from the fact that exp holds in J and that b does not lie in J). In effect: $a^u < 2^{|a|u} < 2^{|a|^2} < 2^a \leq 2b + 1$. Since a is in K, this calculation also shows that exp fails in K.[15] $\dashv$

The previous argument can be adapted to show that, for each n, the theory $\mathsf{I}\Delta_0 + \Omega_n + \neg\,\mathrm{exp}$ is interpretable in Q. For Ω_1 one takes $K(x) :\equiv \exists u^J\, (x \leq a^{|a|^u})$, for Ω_2 one takes $K(x) :\equiv \exists u^J\, (x \leq a^{|a|^{\|a\|^u}})$, etc.

PROPOSITION 9 (Solovay). *There is a sentence ϕ such that both $\mathsf{I}\Delta_0 + \Omega_1 + \phi$ and $\mathsf{I}\Delta_0 + \Omega_1 + \neg\phi$ are theories interpretable in* Q.

PROOF. It is sufficient to find a sentence ϕ such that both $\mathsf{I}\Delta_0 + \Omega_1 + \phi$ and $\mathsf{I}\Delta_0 + \Omega_1 + \neg\phi$ are interpretable in $\mathsf{I}\Delta_0 + \Omega_2 + \neg\,\mathrm{exp}$. Clearly, the sentence ϕ_{even} of the proof of Proposition 8 does the job. $\dashv$

REFERENCES

[1] J. BARWISE, ***Admissible set theory and structures: An approach to definability theory***, Perspectives in Mathematical Logic, Springer, 1975.

[2] S. BELLANTONI and S. COOK, *A new recursion-theoretic characterization of the polytime functions*, ***Computational Complexity***, vol. 2 (1992), pp. 97–110.

[3] J. BENNETT, ***On spectra***, Ph.D. thesis, Princeton University, Princeton, New Jersey, 1962.

[4] J. BURGESS and A. HAZEN, *Predicative logic and formal arithmetic*, ***Notre Dame Journal of Formal Logic***, vol. 39 (1998), pp. 1–17.

[5] S. R. BUSS, ***Bounded arithmetic***, Ph.D. thesis, Princeton University, 1985, a revision of this thesis was published by Bibliopolis (Naples) in 1986.

[6] ———, *Nelson's work on logic and foundations and other reflections on foundations of mathematics*, ***Diffusion, quantum theory, and radically elementary mathematics*** (W. Faris, editor), Princeton University Press, 2006, pp. 183–208.

[7] A. CANTINI, *Asymmetric interpretations for bounded theories*, ***Mathematical Logic Quarterly***, vol. 42 (1996), no. 1, pp. 270–288.

[8] C. DIMITRACOPOULOS, ***Matiyasevič theorem and fragments of arithmetic***, Ph.D. thesis, University of Manchester, 1980.

[9] W. Ewald (editor), ***From Kant to Hilbert***, vol. 2, Oxford University Press, 1996.

[15]The anonymous referee pointed to us that the detour throught $\mathsf{I}\Sigma_2$ can be avoided (see section 6.2.1 of [43]). The gist of the alternative proof is different from the one in the text. It also uses, nevertheless, a version of the Hilbert–Bernays completeness theorem (suitable for bounded theories of arithmetic) that goes by the name of *interpretation existence lemma* (versions of which are theorem 3.1 of [34] or corollary 6.1 of [44]).

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

[11] ———, *Systems of predicative analysis*, ***The Journal of Symbolic Logic***, vol. 29 (1964), no. 1, pp. 1–30.

[12] A. M. Fernandes, ***Investigações em sistemas de análise exequível***, Ph.D. thesis, Universidade de Lisboa, Portugal, 2001, (in Portuguese).

[13] ———, *A new conservation result of* WKL_0 *over* RCA_0, ***Archive for Mathematical Logic***, vol. 41 (2002), pp. 55–63.

[14] ———, *Strict* Π^1_1*-reflection in bounded arithmetic*, ***Archive for Mathematical Logic***, vol. 49 (2010), pp. 17–34.

[15] A. M. Fernandes and F. Ferreira, *Groundwork for weak analysis*, ***The Journal of Symbolic Logic***, vol. 67 (2002), no. 2, pp. 557–578.

[16] F. Ferreira, ***Polynomial time computable arithmetic and conservative extensions***, Ph.D. thesis, Pennsylvania State University, USA, 1988.

[17] ———, *Polynomial time computable arithmetic*, ***Logic and computation*** (W. Sieg, editor), American Mathematical Society, 1990, pp. 161–180.

[18] ———, *A feasible theory for analysis*, ***The Journal of Symbolic Logic***, vol. 59 (1994), no. 3, pp. 1001–1011.

[19] F. Ferreira and G. Ferreira, *The Riemann integral in weak systems of analysis*, ***Journal of Universal Computer Science***, vol. 14 (2008), no. 6, pp. 908–937.

[20] G. Ferreira, ***Aritmética computável em espaço polinomial***, Master's thesis, Universidade de Lisboa, 2001, (in Portuguese).

[21] ———, ***Sistemas de análise fraca para a integração***, Ph.D. thesis, Universidade de Lisboa, 2006, (in Portuguese).

[22] H. Friedman, *FOM*: 73:*Hilbert's program wide open*?, FOM e-mail list: `http://www.cs.nyu.edu/pipermail/fom/1999-December/003511.html`, December 20, 1999.

[23] M. J. Greenberg, *Old and new results in the foundations of elementary plane Euclidean and non-Euclidean geometries*, ***American Mathematical Monthly***, vol. 117 (2010), no. 3, pp. 198–219.

[24] P. Hájek, *Interpretability and fragments of arithmetic*, ***Arithmetic, proof theory and computational complexity*** (P. Clote and J. Krajíček, editors), Clarendon Press, Oxford, 1993, pp. 185–196.

[25] P. Hájek and P. Pudlák, ***Metamathematics of first-order arithmetic***, Perspectives in Mathematical Logic, Springer, 1993.

[26] J. Krajíček, ***Bounded arithmetic, propositional logic, and complexity theory***, Encyclopedia of Mathematics and its Applications, vol. 60, Cambridge University Press, 1995.

[27] D. Marker, ***Model theory: An introduction***, vol. 217, Springer, 2002.

[28] E. Nelson, ***Predicative arithmetic***, Mathematical Notes, Princeton University Press, 1986.

[29] ———, *Confessions of an apostate mathematician*, paper available at `https://web.math.princeton.edu/~nelson/papers.html`, 1995.

[30] ———, *Completed versus incomplete infinity in arithmetic*, paper available at `https://web.math.princeton.edu/~nelson/papers.html`, 2005.

[31] ———, *Warning signs of a possible collapse of contemporary mathematics*, ***Infinity. New research frontiers*** (M. Heller and H. Woodin, editors), Cambridge University Press, 2011, pp. 76–85.

[32] C. Parsons, ***Mathematical thought and its objects***, Cambridge University Press, 2008.

[33] H. Poincaré, *Les mathématiques et la logique*, ***Revue de métaphysic et de moral***, vol. 14 (1906), pp. 294–317, English translation in [9], pp. 1052–1071.

[34] P. Pudlák, *Cuts, consistency statements and interpretations*, ***The Journal of Symbolic Logic***, vol. 50 (1985), no. 2, pp. 423–441.

[35] R. M. Robinson, *An essentially undecidable axiom system*, ***Proceedings of the International Congress of Mathematicians, Cambridge 1950***, vol. 1, American Mathematics Society, 1952, pp. 729–730.

[36] G. Sacks, ***Higher recursion theory***, Perspectives in Mathematical Logic, Springer, 1990.

[37] J. T. Schwartz, *Do the integers exist? The unknowability of arithmetic consistency*, ***Communications on Pure and Applied Mathematics***, vol. 58 (2005), no. 9, pp. 1280–1286.

[38] S. G. Simpson, ***Subsystems of second order arithmetic***, Perspectives in Mathematical Logic, Springer, 1999.

[39] C. Smoryński, *The incompleteness theorems*, ***Handbook of mathematical logic*** (J. Barwise, editor), Studies in Logic and the Foundations of Mathematics, vol. 90, North Holland, 1977, pp. 821–865.

[40] R. M. Solovay, *Letter to P. Hajék*, August 1976.

[41] A. Tarski, *A decision method for elementary algebra and geometry*, Technical report, Rand Corporation, 1957, second edition. Prepared for publication by J. C. C. McKinsey. First edition published in 1948.

[42] A. Tarski, A. Mostowski, and R. Robinson, ***Undecidable theories***, North-Holland, 1953.

[43] A. Visser, *Interpretability logic*, ***Mathematical logic*** (P. P. Petkov, editor), Plenum Press, 1990, pp. 307–359.

[44] ———, *The formalization of interpretability*, ***Studia Logica***, vol. 50 (1991), no. 1, pp. 81–105.

[45] ———, *The unprovability of small inconsistency*, ***Archive for Mathematical Logic***, vol. 32 (1993), no. 4, pp. 275–298.

[46] V. Voevodsky, *What if current foundations of mathematics are inconsistent?*, Video lecture commemorating the 80th anniversary of the Institute for Advanced Study (Princeton), 2010, available at `http://video.ias.edu/voevodsky-80th`.

[47] H. Weyl, ***Das Kontinuum***, Veit, Leipzig, 1918.

[48] A. Wilikie, *On sentences interpretable in systems of arithmetic*, ***Logic colloquium 1984*** (J. Paris, A. Wilkie, and G. Wilmers, editors), North Holland, 1986, pp. 329–342.

[49] A. Wilkie and J. Paris, *On the scheme of induction for bounded arithmetic formulas*, ***Annals of Pure and Applide Logic***, vol. 35 (1987), pp. 261–302.

[50] D. Zambella, *Notes on polynomially bounded arithmetic*, ***The Journal of Symbolic Logic***, vol. 61 (1996), pp. 942–966.

UNIVERSIDADE DE LISBOA, FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA, CAMPO GRANDE, ED. C6
1749-016 LISBOA, PORTUGAL
E-mail: fjferreira@fc.ul.pt

UNIVERSIDADE LUSÓFONA DE HUMANIDADES E TECNOLOGIAS
DEPARTAMENTO DE MATEMÁTICA, AV. DO CAMPO GRANDE 376
1749-024 LISBOA, PORTUGAL
and
UNIVERSIDADE DE LISBOA, FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA, CAMPO GRANDE, ED. C6
1749-016 LISBOA, PORTUGAL
E-mail: gildafer@cii.fc.ul.pt