ON THE OPTIMALITY OF CONSERVATION RESULTS FOR LOCAL REFLECTION IN ARITHMETIC

A. Cordón–Franco and F.F. Lara–Martín

Dpto. Ciencias de la Computación e Inteligencia Artificial
Facultad de Matemáticas. Universidad de Sevilla
C/ Tarfia, s/n, 41012 Sevilla (Spain)
{acordon,fflara}@us.es

November 2, 2011

Abstract. Let $T$ be a recursively enumerable theory extending Elementary Arithmetic $EA$. L.D. Beklemishev proved that the $\Sigma_2$ local reflection principle for $T$, $\text{Rfn}_{\Sigma_2}(T)$, is conservative over the $\Sigma_1$ local reflection principle, $\text{Rfn}_{\Sigma_1}(T)$, with respect to boolean combinations of $\Sigma_1$–sentences; and asked whether this result is best possible. In this work we answer Beklemishev’s question by showing that $\Pi_2$–sentences are not conserved for $T = EA + \text{“}f\text{ is total}\text{”}$, where $f$ is any non-decreasing computable function. We also discuss how this result generalizes to $n > 0$ and obtain as an application that for $n > 0$, $I\Pi_{n+1}$ is conservative over $I\Sigma_n$ with respect to $\Pi_{n+2}$–sentences. Our methods are model–theoretic and we exploit the connections between reflection principles and induction schemes.

1. Introduction

This work was motivated by a question of L.D. Beklemishev on the optimality of a conservation result for reflection principles in first order arithmetic. Reflection principles for a given theory $T$ are axiom schemes expressing the soundness of $T$, i.e. the statement that “if a formula $\varphi$ is provable in $T$ then $\varphi$ is valid.” More precisely, if $T$ is a recursively enumerable (r.e.) arithmetic theory extending Elementary Arithmetic $EA$ and $\Box_T(x) = \exists y \text{Pr}_f T(x, y)$ denotes a standard provability predicate for $T$, the local reflection principle for $T$ is the axiom scheme given by

$$\text{Rfn}(T) : \quad \Box_T(\uparrow \varphi) \to \varphi,$$

where $\varphi$ ranges over all sentences of the language of $T$ and $\uparrow \varphi$ denotes (the numeral of) the Gödel number of $\varphi$. The term local refers to the fact that the scheme

---

Partially supported by grant MTM2008–06435, Ministerio de Educación e Innovación, Spain.
is restricted to sentences, in contrast with the uniform reflection principle for $T$, where formulas with free variables are allowed:

$$RFN(T) : \Box_T (\varphi(x)) \to \varphi(x).$$

Partial reflection principles for $T$ (denoted $Rfn_\Gamma(T)$ or $RFN_\Gamma(T)$, respectively) are obtained from the local or uniform reflection principle for $T$ by imposing a restriction that $\varphi$ may only range over a certain class of formulas $\Gamma$. For $\Gamma$ one typically takes one of the classes $\Sigma_n$ or $\Pi_n$ of the arithmetic hierarchy.

Two interesting topics in the investigation of reflection principles are: i) conservativity, axiomatizability and other proof-theoretic properties of theories axiomatized by these principles, and ii) results à la Kreisel–Lévy [14] establishing the equivalence between reflection principles and different forms of induction in arithmetic. Here we contribute some new results of both types.

In the study of conservation results for local reflection principles provability logic has played a prominent role. In fact, using provability logic techniques in [3] L.D. Beklemishev showed that the local reflection hierarchy enjoys the peculiar conservation property that over $T$, full local reflection $Rfn(T)$ is $\Gamma$–conservative over local reflection restricted to $\Gamma$–sentences $Rfn_\Gamma(T)$ for $\Gamma = \Sigma_n$ or $\Pi_n$. More precisely,

**Theorem 1.1 (Beklemishev, [3]).**

1. For $n > 1$, $T + Rfn(T)$ is conservative over $T + Rfn_{\Sigma_n}(T)$ with respect to $\Sigma_n$–sentences (and dually for $\Pi_n$).
2. $T + Rfn(T)$ is conservative over $T + Rfn_{\Sigma_1}(T)$ with respect to $B(\Sigma_1)$ (= boolean combinations of $\Sigma_1$) sentences.

In [3] Beklemishev also noted that part (1) in the above theorem is best possible with respect to arithmetic complexity. Later, in his survey paper on reflection principles in arithmetic [7], Beklemishev raised the problem of determining the optimality of part (2). It is clear that, in general, $T + Rfn(T)$ is neither $\Pi_2$ nor $\Sigma_2$ conservative over $T + Rfn_{\Sigma_1}(T)$, for otherwise $T + Rfn_{\Pi_2}(T)$ (resp. $T + Rfn_{\Sigma_2}(T)$) would collapse to $T + Rfn_{\Sigma_1}(T)$ and it is known that the local reflection hierarchy is proper unless its $\Sigma_1$ level is already inconsistent. Nevertheless, it immediately follows from part (2) that both $T + Rfn_{\Pi_2}(T)$ and $T + Rfn_{\Sigma_2}(T)$ are $B(\Sigma_1)$–conservative over $T + Rfn_{\Sigma_1}(T)$ and now there is room for improvement since raising these conservation results would not imply a collapse of the hierarchy. In fact, it follows from lemma 4.35 in [7] that conservativity of $T + Rfn_{\Pi_2}(T)$ over $T + Rfn_{\Sigma_1}(T)$ can be extended to $\Sigma_2$–sentences if $T \subseteq \Pi_2$. In contrast, the corresponding question for $T + Rfn_{\Sigma_2}(T)$ seems to be open.

**Question 1 (Beklemishev).** Is $T + Rfn_{\Sigma_1}(T)$ conservative over $T + Rfn_{\Sigma_1}(T)$ with respect to $\Pi_2$–sentences?
Note that, for theories $T$ of arithmetical complexity $B(\Sigma_1)$, a positive answer to Question 1 would imply the equivalence between the $\Pi_2$–consequences and the $B(\Sigma_1)$–consequences of $T + \text{Rfn}_{\Sigma_2}(T)$. In general, the $\Pi_2$–consequences of an arithmetic theory are much stronger than its $B(\Sigma_1)$–consequences, the totality of a computable function with more than elementary growth being a natural $\Pi_2$–axiom separating both sets of sentences. Nevertheless, for $T + \text{Rfn}_{\Sigma_2}(T)$ we have the special situation that it has arithmetical complexity $\Sigma_2$. As a result, $T + \text{Rfn}_{\Sigma_2}(T)$ and its $B(\Sigma_1)$–consequences share the same class of provably total computable functions; and the previous rate of growth separation argument does not work. This motivates the following problem on general arithmetic theories, of which Beklemishev’s Question 1 for $T \subseteq B(\Sigma_1)$ is a particular case (as usual, $\text{Th}_\Gamma(S)$ denotes the $\Gamma$–consequences of a theory $S$):

**Question 2.** Suppose $S$ contains $EA$. Are $\text{Th}_{\Pi_2}(\text{Th}_{\Sigma_2}(S))$ and $\text{Th}_{B(\Sigma_1)}(S)$ deductively equivalent?

In this work we solve in the negative both questions and apply the proof ideas to obtain some new results on local reflection principles. More precisely, the main results of the paper are:

- *(Answer to Question 1)* If $T = EA + \text{“} f \text{ is total} \text{“}$, where $f$ is any non-decreasing computable function, then $T + \text{Rfn}_{\Sigma_2}(T)$ is not $\Pi_2$–conservative over $T + \text{Rfn}_{\Sigma_1}(T)$. Thus, known $B(\Sigma_1)$–conservativity between these theories is best possible.

- *(Answer to Question 2)* If $S$ is a consistent, r.e. extension of $EA + \text{Rfn}_{\Sigma_2}(EA)$, then $\text{Th}_{\Pi_2}(\text{Th}_{\Sigma_2}(S))$ is strictly stronger than $\text{Th}_{B(\Sigma_1)}(S)$.

- *(Results à la Kreisel–Lévy)* We characterize the $\Pi_2$ and the $B(\Sigma_1)$ consequences of $T + \text{Rfn}_{\Sigma_2}(T)$ in terms of certain variants of the $\Sigma_1$–induction rule. The ideas involved yield nice applications. Firstly, we show that $T + \text{Rfn}_{\Sigma_1}(T)$ is equivalent to a parameter free version of the $\Pi_1$–induction rule $\Pi_{1}^{\text{IR}}$. Secondly, we point out that, modulo $EA$, the $\omega$ times iterated consistency assertion for a theory $T$ corresponds to the $\Pi_1$–induction rule (this result was known only modulo a base theory strictly stronger than $EA$).

- *(Generalizations to $n > 0$)* This involves considering relativized local reflection $\text{Rfn}_{\Sigma_2}^{n}(T)$. We prove that $T + \text{Rfn}_{\Sigma_2}^{n}(T)$ is not $\Pi_{n+2}$–conservative over $T + \text{Rfn}_{\Sigma_{n+1}}^{n}(T)$ for $B(\Sigma_{n+1})$–extensions of $EA$; as well as we show the positive result that $T + \text{Rfn}_{\Sigma_{n+2}}^{n}(T)$ is $\Pi_{n+2}$–conservative over $T + \text{Rfn}_{\Sigma_{n+1}}^{n}(T)$. These results have interesting applications to the study of the relative strength of the scheme of parameter free $\Pi_{n+1}$–induction $\Pi_{n+1}^{\text{IR}}$.

Although at first glance the answer to Question 2 might seem to be rather obvious, it should be compared with the following result stating that a dual version of that question has a positive answer.

3
**Proposition 1.2.** If $S$ implies $EA$ then $Th_{\Sigma_2}(Th_{\Pi_2}(S)) \equiv Th_{B(\Sigma_1)}(S)$.

**Proof.** It is obvious that $Th_{B(\Sigma_1)}(S) \subseteq Th_{\Sigma_2}(Th_{\Pi_2}(S))$. For the opposite direction, assume that $\mathfrak{A}$ is a model of $Th_{B(\Sigma_1)}(S)$. Consider the theory

$$S' = Th_{\Sigma_1}(\mathfrak{A}) + Th_{\Pi_1}(\mathfrak{A}) + S,$$

where $Th_{\Gamma}(\mathfrak{A})$ denotes the set of all $\Gamma$-sentences which are true in the model $\mathfrak{A}$. Clearly, $S'$ is consistent. Let $\mathfrak{B}$ be a model of $S'$. We denote by $K_1(\mathfrak{A})$ (resp. $K_1(\mathfrak{B})$) the submodel consisting of all $\Sigma_1$ definable elements of $\mathfrak{A}$ (resp. $\mathfrak{B}$). By using $\Delta_0$-minimization (which is available thanks to $EA$), we obtain that $K_1(\mathfrak{A})$ and $K_1(\mathfrak{B})$ are $\Sigma_1$-elementary substructures. Thus, $K_1(\mathfrak{A}) \models Th_{\Pi_2}(S)$ and $\mathfrak{A} \models Th_{\Sigma_2}(K_1(\mathfrak{A}))$.

But it follows from the fact that $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same $\Sigma_1$-sentences that $K_1(\mathfrak{A})$ and $K_1(\mathfrak{B})$ are isomorphic (if $\varphi(x)$ defines $a \in \mathfrak{A}$, map $a$ to the unique $b \in \mathfrak{B}$ satisfying $\varphi(x)$ in $\mathfrak{B}$). Consequently, $\mathfrak{A} \models Th_{\Sigma_2}(Th_{\Pi_2}(S))$. 

**Remark 1.3.**

1. Note that the presence of the exponential function in the language of $S$ is irrelevant to the proof of Proposition 1.2. In fact, it works for theories $S$ in any language $L$ extending that of Peano Arithmetic $\{0, 1, +, \cdot, \leq\}$ as long as $S$ contains the scheme of induction for $\Delta_0(L)$-formulas. In addition, the result easily generalizes to an arbitrary $n > 0$, namely, $Th_{\Sigma_{n+2}}(Th_{\Pi_{n+2}}(S)) \equiv Th_{B(\Sigma_{n+1})}(S)$ if $S$ contains the scheme of parameter free $\Sigma_n$-induction $I\Sigma_n^-$.

2. Despite its simplicity it seems that Proposition 1.2 has not been known, even though some particular cases have been observed in the literature. Firstly, in [13] the authors described it as surprising the fact that, for $n > 0$, the $B(\Sigma_{n+1})$-consequences of $I\Sigma_n$ form an axiomatization of the $\Sigma_{n+2}$-consequences of $I\Sigma_n$. We know from Proposition 1.2 that this is a general property of every $\Pi_{n+2}$-extension of $I\Sigma_n^-$. Secondly, in [6] the author proved that over $EA$, $Rfn_{\Sigma_1}(T)$ is $\Sigma_2$-conservative over $Rfn_{\Sigma_1}(T)$ and remarked that this result can be considered unexpected in that it implies that the $\Sigma_2$-consequences of $Rfn_{\Sigma_1}(T)$ can be axiomatized by a set of formulas of lower arithmetical complexity: $Rfn_{\Sigma_1}(T) \subseteq B(\Sigma_1)$. Again, it follows from Proposition 1.2 that this fact is true for every $\Pi_2$-theory extending $\Delta_0$-induction.

3. Proposition 1.2 provides us with a recipe for rising a $B(\Sigma_1)$-conservation result to $\Sigma_2$-conservation. Namely, if $T_1$ is a $\Pi_2$-theory extending $\Delta_0$-induction and $T_1$ is $B(\Sigma_1)$-conservative over $T_2$, then $\Sigma_2$-sentences are also conserved. In particular, this gives us a simple proof of the known fact that the dual version of Question 1 has a positive answer for $\Pi_2$-theories, i.e., $T + Rfn_{\Pi_2}(T)$ is $\Sigma_2$-conservative over $T + Rfn_{\Sigma_1}(T)$ if $T \subseteq \Pi_2$. 

4
Thus, our work also brings into evidence the different behaviour of the dual classes $Th_{\Pi_1}(Th_{\Sigma_2}(S))$ and $Th_{\Sigma_2}(Th_{\Pi_1}(S))$ for a general arithmetic theory $S$.

Our methods are model–theoretic and we exploit the connections between reflection principles and induction schemes. Actually, it is a theorem of Beklemishev that over $EA$, $Rfn_{\Sigma_2}(T)$ contains the scheme of parameter free $\Pi_1$–induction

$$I\Pi_1^- : \varphi(0) \land \forall x (\varphi(x) \rightarrow \varphi(x + 1)) \rightarrow \forall x \varphi(x),$$

where $\varphi(x) \in \Pi_1^-$ (as usual, we write $\varphi(x) \in \Gamma^-$ to mean that $\varphi$ is in $\Gamma$ and contains no other free variables than $x$). In Section 2 of the present paper we develop a model–theoretic analysis of $I\Pi_1^-$ and obtain the somewhat surprising result that $I\Pi_1^-$ (and then also $Rfn_{\Sigma_2}(T)$) allows for certain variants of the $\Sigma_1$–induction scheme, that we call local $\Sigma_1$–induction. These variants are obtained by restricting the induction scheme to definable elements and their use is the main tool for obtaining our results in the subsequent sections.

Our notation is standard and we assume that the reader is familiar with the basic notions of first order arithmetic (we recommend [12] and [11] for a detailed introduction to the subject; [13] for results on parameter free schemes; and [7] for information on reflection principles). To a large extent, our results are independent of the language we are working in. Nevertheless, for the sake of definiteness we assume that we work in the language $\mathcal{L}_\text{exp} = \{0, 1, +, \cdot, <, \text{exp}\}$ extending that of Peano Arithmetic $PA$ with a symbol for the function $2^x$. Also, we assume that all the theories we shall deal with are extensions of Elementary Arithmetic $EA$, which is a first order theory axiomatized by a finite set of defining axioms for the symbols in $\mathcal{L}_\text{exp}$ plus the scheme of induction for bounded ($\Delta_0$) formulas of $\mathcal{L}_\text{exp}$. Finally, we also assume, sometimes without explicit mention, that every theory $T$ for which we consider reflection principles is an elementary presented extension of $EA$; the set of its axioms being represented by an elementary formula of the form $Ax_{EA}(x) \lor Ax_T(x)$.

## 2. On Local Induction

Two natural ways to weaken the induction axiom for a formula $\varphi(x, v)$ are: i) to limit the use of parameters $v$, and ii) to restrict the elements $x$ for which the axiom claims $\varphi(x, v)$ to hold. In the present paper we use the term local induction to mean those axiom schemes obtained from the usual induction scheme by imposing these two types of restrictions. More precisely, suppose $A$ and $B$ each define a subset of the universe of a model and $\Gamma$ is a class of formulas.

- The scheme of parameter free $\Gamma$–induction up to $A$, denoted $I(\Gamma^- , A)$, is given by:

$$\varphi(0) \land \forall x (\varphi(x) \rightarrow \varphi(x + 1)) \rightarrow \forall x \in A^\Gamma \varphi(x),$$
where \( \varphi(x) \in \Gamma^\sim \).

- The scheme of \( \Gamma \)-induction up to \( A \) with parameters in \( B \), denoted \( I(\Gamma, A, B) \), is given by:
  \[
  \forall v \in B \left[ \varphi(0, v) \land \forall x \left( \varphi(x, v) \rightarrow \varphi(x + 1, v) \right) \rightarrow \forall x \in A \varphi(x, v) \right],
  \]
  where \( \varphi(x, v) \in \Gamma \).

Note that, for the previous axiom schemes to be actual theories in the language of arithmetic, it is not necessary to assume \( A \) and \( B \) to be definable subsets. Indeed, it is sufficient that \( A \) and \( B \) can be defined by a scheme of formulas. This can be easily achieved when \( A \) and \( B \) are taken to be substructures of definable elements.

We say that \( a \) is a \( \Sigma_n \)-definable element of a model \( \mathfrak{A} \) if there is \( \varphi(x) \in \Sigma_n \) such that \( a \) is the unique element satisfying \( \varphi(x) \) in \( \mathfrak{A} \). In this work we shall deal with the following substructures of definable elements.

- \( K_n(\mathfrak{A}) \) denotes the set of all \( \Sigma_n \)-definable elements of \( \mathfrak{A} \); \( I_n(\mathfrak{A}) \) denotes the initial segment of \( \mathfrak{A} \) determined by \( K_n(\mathfrak{A}) \).
- \( K^1_n(\mathfrak{A}) = K_n(\mathfrak{A}, I_n(\mathfrak{A})) \) denotes the set of all elements of \( \mathfrak{A} \) which are \( \Sigma_n \)-definable with a parameter from \( I_n(\mathfrak{A}) \); \( I^1_n(\mathfrak{A}) \) denotes the initial segment of \( \mathfrak{A} \) determined by \( K^1_n(\mathfrak{A}) \).

The key observation is that membership to these substructures can be expressed, in a canonical way, in the language of arithmetic. Given a formula \( \delta(x, v) \), we write \( Def_\delta(x, v) \) to denote \( \delta(x, v) \land \forall x, x' (\delta(x, v) \land \delta(x', v) \rightarrow x = x') \); or we simply write \( Def_\delta(x) \) if \( \delta(x) \) does not contain any parameters. Then, a quantifier of the form \( \forall x \in K_n \)” in front of a formula \( \Psi(x) \) is intended as a shorthand for the scheme:

\[
\{ \forall x [Def_\delta(x) \rightarrow \Psi(x)] : \delta \in \Sigma_n \}.
\]

Likewise, \( \forall x \in I_n \Psi(x) \)” unravels to

\[
\{ \forall x, y [Def_\delta(y) \land x \leq y \rightarrow \Psi(x)] : \delta \in \Sigma_n \},
\]

a quantifier of the form \( \forall x \in K^1_n \Psi(x) \)” unravels to

\[
\{ \forall x, y, z [Def_\delta_1(z) \land y \leq z \land Def_\delta_2(x, y) \rightarrow \Psi(x)] : \delta_1, \delta_2 \in \Sigma_n \},
\]

and, finally, \( \forall x \in I^1_n \Psi(x) \)” unravels to

\[
\{ \forall x, y, z, u [Def_\delta_1(u) \land z \leq u \land Def_\delta_2(y, z) \land x \leq y \rightarrow \Psi(x)] : \delta_1, \delta_2 \in \Sigma_n \}.
\]

Thus, \( I(\Sigma^1_\Gamma, K^1) \), \( I(\Sigma^1_\Gamma, I^1) \), \( I(\Sigma^1_\Gamma, I, I^1) \) and \( I(\Sigma^1, I, I^1) \) are to be seen as axiom schemes in the language of arithmetic. Our first result says that, although \( III^1_\Gamma \) is much weaker than the \( \Sigma^1_\Gamma \)-induction scheme \( I\Sigma^1_\Gamma \), \( III^1_\Gamma \) allows for (and, in fact, equivalent to) the variants of local \( \Sigma^1_\Gamma \)-induction up to \( \Sigma^1 \)-definable elements listed above. The proof puts together a number of now well--established arguments for analyzing parameter free schemes that appeared originally in [13].
Proposition 2.1. Over $EA$, the following theories are equivalent.

(1) $\Pi^1_1$.
(2) $I(\Sigma^*_1, K_1)$.
(3) $I(\Sigma_1, I_1, K^*_1)$.

Proof. (1$\Rightarrow$2): Assume $\mathfrak{A} \models I\Pi^1_1$ and $\mathfrak{A} \models \varphi(0) \land \forall x (\varphi(x) \rightarrow \varphi(x + 1)) \land \neg \varphi(a)$, where $\varphi(x) \in \Sigma^*_1$ and $a \in K_1(\mathfrak{A})$. Consider $\delta(z) \in \Sigma_1$ defining $a$ in $\mathfrak{A}$ and put
\[ \theta(x) \equiv \forall z (\delta(z) \rightarrow \neg \varphi(z - x)). \]
Clearly, $\theta(x) \in \Pi^1_1$ and $\mathfrak{A} \models \theta(a)$ and so $\mathfrak{A} \models \neg \varphi(0)$, which is a contradiction.

(2$\Rightarrow$3): Assume $\mathfrak{A} \models I(\Sigma^*_1, K_1)$ and that $\mathfrak{A}$ satisfies the antecedent of the induction axiom for $\varphi(x, b)$, where $\varphi(x, v) \in \Sigma_1$ and $b \in K^*_1(\mathfrak{A})$. Consider $\alpha \in I_1(\mathfrak{A})$. We must show $\mathfrak{A} \models \varphi(a, b)$. To this end, pick $b' \in I_1(\mathfrak{A})$ such that $\delta(w, b') \in \Sigma_1$ defines $b$ and pick $c \in K_1(\mathfrak{A})$ such that $a, b' \leq c$. Put
\[ \theta(s, c) \equiv \forall x, v \leq c [\exists w (\delta(w, v) \land \varphi(x, w)) \rightarrow (s)_{(x,v)} \neq 0], \]
where $(s)_i$ denotes the $i$–th component of the sequence coded by $s$ and $(\cdot, \cdot)$ denotes the Cantor pairing function. Clearly, $\theta(s, v) \in \Pi_1$ and there is $d \in K_1(\mathfrak{A})$ with $\mathfrak{A} \models \theta(d, c)$ (take $d$ to be the code of a sequence of $(c, c)$–many 1’s). Then we have

Claim 2.1.1. There exists the least $s$ satisfying $\theta(s, c)$ in $\mathfrak{A}$, say $e$.

Proof. Towards a contradiction, assume not. Consider $\delta'(v) \in \Sigma_1$ defining $c$ in $\mathfrak{A}$ and write $\theta(s, v) \equiv \forall y \theta_0(s, y, v)$, with $\theta_0 \in \Delta_0$. Put
\[ \psi(x) \equiv \exists v (\delta'(v) \land \exists s \leq x \neg \theta_0(s, (t)s, v)). \]
Clearly, $\psi(x) \in \Sigma^*_1$ and $\mathfrak{A} \models \psi(x) \rightarrow \forall s \leq x \neg \theta(s, c)$. The antecedent of the induction axiom for $\psi$ is valid in $\mathfrak{A}$, for otherwise there would be a least $s$ satisfying $\theta(s, c)$. So, $\mathfrak{A} \models \psi(d)$ by $I(\Sigma^*_1, K_1)$ and hence $\mathfrak{A} \models \neg \theta(d, c)$, which gives us the desired contradiction.

It follows from the minimality of $e$ that $\mathfrak{A} \models \forall x \leq c (\varphi(x, b) \rightarrow (e)_{(x,b)} \neq 0)$. Thus, $\varphi(a, b)$ follows by applying $\Delta_0$–induction to $x \leq a \rightarrow (e)_{(x,b)} \neq 0$.

(3$\Rightarrow$1): Assume $\mathfrak{A} \models I(\Sigma_1, I_1, K^*_1)$ and $\mathfrak{A} \models \varphi(0) \land \forall x (\varphi(x) \rightarrow \varphi(x + 1)) \land \exists x \neg \varphi(x)$, with $\varphi(x) \in \Pi^1_1$. Recall that $K_1(\mathfrak{A})$ is a $\Sigma_1$–elementary substructure of $\mathfrak{A}$, written $K_1(\mathfrak{A}) \prec \mathfrak{A}$. So, there is $a \in K_1(\mathfrak{A})$ such that $\mathfrak{A} \models \neg \varphi(a)$. Put $\theta(x, a) \equiv \neg \varphi(a - x)$. By applying $\Sigma_1$–induction up to $a$ to $\theta(x, a)$, we get that $\mathfrak{A} \models \theta(a, a)$ and, so, $\mathfrak{A} \models \neg \varphi(0)$, which is a contradiction.
Remark 2.2.

(1) It immediately follows that over $EA$, the following theories are all equivalent: $\Pi^+_1$, $I(\Sigma^+_1, \mathcal{K}_1)$, $I(\Sigma^+_1, \mathcal{I}_1)$, $I(\Sigma_1, \mathcal{I}_1)$, $I(\Sigma_1, \mathcal{K}_1)$.

(2) The use of exponentiation in the previous proof is unessential. A similar result holds for $\Pi^+_1$ formulated in the language \{0, 1, +, ·, <\} over the scheme of bounded induction $I\Delta_0$. In addition, the result easily generalizes to $\Pi^{n+1}_1$, $I(\Sigma^{n+1}_1, \mathcal{K}_{n+1})$ and $I(\Sigma^{n+1}_1, \mathcal{I}_{n+1}, \mathcal{K}^1_{n+1})$ for $n > 0$ over the scheme of parameter free $\Sigma_n$–induction $I\Sigma^+_n$.

(3) The previous result is best possible in the sense that $\Pi^+_1$ implies neither $I(\Sigma_1, \mathcal{I}_1)$ nor $I(\Sigma^-_1, \mathcal{K}^1_1)$ (for proofs see proposition 2.6 of [8]).

Since $T + Rfn_{\Sigma^2}(T)$ contains $\Pi^+_1$ by theorem 1 of [5], it immediately follows that

**Proposition 2.3.** $T + Rfn_{\Sigma^2}(T)$ implies $I(\Sigma_1, \mathcal{I}_1, \mathcal{K}^1_1)$.

The fact that $Rfn_{\Sigma^2}(T)$ allows for local $\Sigma_1$–induction up to $\Sigma_1$–definable elements will be important in the next sections to prove the main results of the paper. Here, as a first application we obtain conservation results for: (1) the scheme of $\Sigma_1$–collection, $B\Sigma_1$; and (2) the closure of a theory $T$ under first order logic and non-nested applications of the $\Sigma_1$–induction rule, $[T, \Sigma_1$–IR$]$.

**Proposition 2.4.** Suppose $T \subseteq \Pi_2$.

(1) $T + B\Sigma_1$ is $\Sigma_2$–conservative over $T + Rfn_{\Sigma^2}(T)$.

(2) $[T, \Sigma_1$–IR$]$ is $\Sigma_2$–conservative over $T + Rfn_{\Sigma^2}(T)$.

**Proof.** (1): Towards a contradiction, assume $T + B\Sigma_1 \vdash \varphi$ and $\mathfrak{A}$ is a model of $T + Rfn_{\Sigma^2}(T) + \neg \varphi$, with $\varphi$ a $\Sigma_2$–sentence. By considering an elementary extension of $\mathfrak{A}$ if necessary, we may assume that $\mathcal{K}_1(\mathfrak{A})$ is bounded above in $\mathfrak{A}$. Then, $\mathcal{I}_1(\mathfrak{A}) \models B\Sigma_1$ since $\mathcal{I}_1(\mathfrak{A})$ is a proper initial segment of $\mathfrak{A}$. In addition, we have:

**Claim 2.4.1.** $\mathcal{I}_1(\mathfrak{A}) \models Th_{\Pi_2}(\mathfrak{A})$.

**Proof.** Suppose $\mathfrak{A} \models \forall x \exists y \theta(x, y)$, with $\theta(x, y) \in \Delta_0$, and $a \in \mathcal{I}_1(\mathfrak{A})$. Pick $b \in K_1(\mathfrak{A})$ such that $a \leq b$. Define $\delta(z)$ to be $\exists u \forall x \leq z \exists y \leq u \theta(x, y)$. It is clear that $\delta(z) \in \Sigma^+_1$ and $\mathfrak{A} \models \delta(0) \land \forall z (\delta(z) \rightarrow \delta(z + 1))$. Since $T + Rfn_{\Sigma^2}(T) \vdash I(\Sigma_1, \mathcal{I}_1, \mathcal{I}_1)$ by Proposition 2.3, $\mathfrak{A} \models \delta(b)$. Since $K_1(\mathfrak{A}) \prec \mathfrak{A}$ and $K_1(\mathfrak{A}) \prec \mathcal{I}_1(\mathfrak{A})$, $\delta(b)$ is valid in $\mathcal{I}_1(\mathfrak{A})$ too. So, $\mathcal{I}_1(\mathfrak{A}) \models \forall x \leq b \exists y \theta(x, y)$ and hence $\mathcal{I}_1(\mathfrak{A}) \models \exists y \theta(a, y)$. \qed

Consequently, $\mathcal{I}_1(\mathfrak{A})$ satisfies $T + B\Sigma_1 + \neg \varphi$, which gives us the desired contradiction.
We say that a formula \( y \) is EA if

\[ \text{Suppose} \]

Part (2) can also be inferred from known results (firstly, by theorem 2.5 of [2] and secondly, by proposition 4.6 of [6])

\[ \text{To this end, assume } [T, \Sigma_1 – \text{IR}] \vdash \varphi \text{ and } \mathfrak{A} \models T + \text{Rfn}_2(T), \text{ with } \varphi \in \mathcal{B}(\Sigma_1). \text{ We need to show } \mathfrak{A} \models \varphi. \]

Claim 2.4.2. \( K_1(\mathfrak{A}) \models [T, \Sigma_1 – \text{IR}]. \)

Proof. Suppose \( T \vdash \theta(0, \nu) \land \forall x (\theta(x, v) \rightarrow \theta(x + 1, v)), \) with \( \theta(x, v) \in \Sigma_1. \) Consider \( a, b \in K_1(\mathfrak{A}). \) By applying \( \Sigma_1 \)-induction up to \( a \) to \( \theta(x, b) \) (available in \( \mathfrak{A} \) thanks to Proposition 2.3), we get \( \mathfrak{A} \models \theta(a, b). \) Since \( K_1(\mathfrak{A}) \prec_1 \mathfrak{A}, \theta(a, b) \) is valid in \( K_1(\mathfrak{A}) \) too. So, \( K_1(\mathfrak{A}) \models \forall \nu \forall x \theta(x, v). \)

3. An Unboundedness Theorem for \( \text{Th}_{\Sigma_2}(T + \text{Rfn}_2(T)) \)

Suppose \( T = EA + \text{"f is total," where } f \text{ is a computable function. Equipped with } I(\Sigma_1, \mathcal{I}, K_1), \) a candidate for "a hardest problem" in \( \text{Th}_{\Sigma_2}(T + \text{Rfn}_2(T)) \) naturally appears:

\[ \forall x \in \mathcal{I} \forall z \in K_1 \text{ "f}^x(z) \text{ exists,"} \]

where \( f^x \) is the \( x \)-th iterate of \( f. \) In this section we shall show that this is actually the case whenever \( f \) is a nondecreasing computable function with an elementary graph. As a consequence, we shall obtain a so-called unboundedness theorem for \( \text{Th}_{\Sigma_2}(T + \text{Rfn}_2(T)) \) which, in turn, will allow us to answer Questions 1 and 2.

Definition 3.1. We say that a formula \( y = f(x) \) is EA–honest if

\[ \begin{align*}
(1) & \quad y = f(x) \text{ is } \Delta_0, \\
(2) & \quad EA \vdash y = f(x) \rightarrow y \geq 2^x, \text{ and} \\
(3) & \quad EA \vdash x_1 \leq x_2 \land y_1 = f(x_1) \land y_2 = f(x_2) \rightarrow y_1 \leq y_2.
\end{align*} \]

Remark 3.2. Note that a theory \( T \) can be written as \( EA + \text{"f is total,} \) where \( y = f(x) \) is EA–honest, if and only if \( T \) is a finite \( \Pi_2 \)-extension of \( EA \) closed under the \( \Sigma_1 \)-collection rule

\[ \Sigma_1-\text{CR}: \quad \frac{\forall x \exists y \varphi(x, y)}{\forall z \exists u \forall x \leq z \exists y \leq u \varphi(x, y)}. \]
where $\varphi(x,y) \in \Sigma_1$ (for a proof see, e.g., proposition 5.4 of [2]).

Let $y = f(x)$ be an $EA$–honest formula. The iteration of $f$, denoted $y = It_f(x,z)$, is the following $\Delta_0$ formula (that uses elementary coding of sequences):

$\exists s \leq bt$ \text{length}(s) = x+1 \land (s)_0 = z \land \forall i < x ((s)_{i+1} = f((s)_i)) \land y = (s)_x$,

where $bt$ is a bounding term for the code of a sequence consisting of $(x+1)$–many $y$’s. Usually, we shall use the more suggestive notation $y = f^x(z)$ in lieu of $y = It_f(x,z)$.

**Lemma 3.3.** Suppose $y = f(x)$ is $EA$–honest. Then

$$\Pi_1^1 + \forall x \exists! y (y = f(x)) \vdash \forall x \in I_1 \forall z \in I_1^1 \exists y (y = f^x(z)).$$

**Proof.** Since $\Pi_1^1$ contains $I(\Sigma_1, I_1, K_1^1)$ by Proposition 2.1, $\Pi_1^1$ proves

$$\forall x \in I_1 \forall z \in K_1^1 \exists y (y = f^x(z)).$$

But it follows from the fact that $y = f(x)$ defines a nondecreasing function that $EA \vdash z' \leq z \land \exists y (y = f^x(z)) \rightarrow \exists y (y = f^x(z'))$.

Since a quantifier of the form “$\forall x \in I_1$” or “$\forall z \in I_1^1$” in front of a $\Sigma_1$–formula unravels to a scheme of formulas of arithmetical complexity $\forall B(\Sigma_1)$, it immediately follows from Lemma 3.3 that

**Proposition 3.4.** Suppose $y = f(x)$ is $EA$–honest and $T \equiv EA + \forall x \exists y (y = f(x))$. Then, $Th_{\Pi_2}(T + Rfn_{\Sigma_2}(T)) \vdash \forall x \in I_1 \forall z \in I_1^1 \exists y (y = f^x(z)).$

The so–called unboundedness theorems due to Kreisel and Lévy [14] state that $Rfn_{\Pi_1}(T)$ is not contained in any consistent finite extension of $T$ of complexity $\Sigma_n$ (and dually for $Rfn_{\Sigma_n}(T)$). Here we obtain the following variant of these results.

**Theorem 3.5 (Unboundedness).** Suppose $T$ is a finite $\Pi_2$–extension of $EA$ closed under $\Sigma_1$–CR. Then, $Th_{\Pi_2}(T + Rfn_{\Sigma_2}(T))$ is not contained in any consistent, r.e. extension of $T$ by $B(\Sigma_1)$–sentences.

**Proof.** Put $T \equiv EA + \forall x \exists y (y = f(x))$, where $y = f(x)$ is $EA$–honest. Let $\Gamma$ be an r.e. set of $B(\Sigma_1)$–sentences satisfying that $T + \Gamma$ is consistent. We shall construct a model of $T + \Gamma$ in which $Th_{\Pi_2}(T + Rfn_{\Sigma_2}(T))$ fails. First of all, note that $T + \Gamma$ does not imply the set of all true $\Pi_1$ sentences $Th_{\Pi_1}(\mathbb{N})$, for otherwise it would follow that $Th_{\Pi_1}(T + \Gamma) = Th_{\Pi_1}(\mathbb{N})$ and this is impossible since the first set is r.e. and the second one is $\Pi_1^1$–complete. Thus, there is $\mathfrak{A} \models T + \Gamma$ in which $Th_{\Pi_1}(\mathbb{N})$ fails and so $K_1(\mathfrak{A})$ is nonstandard. By considering an elementary extension of $\mathfrak{A}$ if necessary, we may also assume that $K_1(\mathfrak{A})$ is bounded above in $\mathfrak{A}$. Then, it holds that $K_1^1(\mathfrak{A}) \neq I_1(\mathfrak{A})$, for otherwise $I_1(\mathfrak{A})$ would have a proper $\Sigma_1$–elementary end.
extension and hence $I_1(\mathfrak{A}) \models B\Sigma_2$ by theorem B of [17]. But $I\Sigma_1 \subseteq B\Sigma_2$ and $I\Sigma_1$ is well-known to fail in $I_1(\mathfrak{A})$ whenever $I_1(\mathfrak{A})$ is nonstandard. It thus follows that there is $a \in \mathcal{K}_1^1(\mathfrak{A})$ such that $I_1(\mathfrak{A}) < a$. Consider $\varphi(x, y, v) \in \Delta_0$ and $b \in I_1(\mathfrak{A})$ such that $\exists y \varphi(x, y, b)$ defines $a$ in $\mathfrak{A}$. Since $\varphi$ is bounded, there is a minimal $c$ satisfying $\varphi((z)_0, (z)_1, b)$ in $\mathfrak{A}$. It is clear that $(c)_0 = a$ and so $I_1(\mathfrak{A}) < a < c$. Now define $\mathfrak{B}$ to be initial segment of $\mathfrak{A}$ determined by the standard iterations $f^k(c)$, i.e.

$$\mathfrak{B} = \{d \in \mathfrak{A} : \exists k \in \omega, \ A \models \exists y (y = f^k(c) \land d \leq y)\}.$$ It follows from $f(x) \geq 2^x$ that $\mathfrak{B}$ is closed under the function $2^x$ and then it is also closed under sums and products. Consequently, $\mathfrak{B}$ is an initial substructure of $\mathfrak{A}$, $\mathfrak{B} \preceq_0 \mathfrak{A}$, and $\mathcal{K}_1^1(\mathfrak{A}) \subseteq \mathfrak{B}$. In addition, we have

**Claim 3.5.1.** $\mathfrak{B} \models T + \Gamma$.

**Proof.** To see $\mathfrak{B} \models T$, it is sufficient to verify that $y = f(x)$ defines a total function in $\mathfrak{B}$. Consider $e \in \mathfrak{B}$. By the definition of $\mathfrak{B}$, there is $k \in \omega$ such that $e \leq f^k(c)$. Since $y = f(x)$ is $EA$-honest, $\mathfrak{A} \models f(e) \leq f(f(k(c))) = f^{k+1}(c)$. So, $f(c) \in \mathfrak{B}$, as required. To see $\mathfrak{B} \models \Gamma$, consider $\theta \in \Gamma$. By logical operations,

$$\theta = (\theta_1^0 \lor \exists x \theta_2^0(x)) \land \ldots \land (\theta_n^0 \lor \exists x \theta_2^n(x)),$$

with $\theta_1^i \in \Pi_1$ and $\theta_2^i(x) \in \Delta_0$. Fix $i \leq n$. Since $\mathfrak{A}$ satisfies $\Gamma$, $\mathfrak{A} \models \theta_1^i \lor \exists x \theta_2^i(x)$.

**Case 1:** $\mathfrak{A} \models \theta_1^i$. Then $\theta_1^i$ is also valid in $\mathfrak{B}$ since $\mathfrak{B} \preceq_0 \mathfrak{A}$.

**Case 2:** $\mathfrak{A} \models \exists x \theta_2^i(x)$. Then there is $a_i \in \mathcal{K}_1^1(\mathfrak{A})$ such that $\mathfrak{A} \models \theta_2^i(a_i)$. But $a_i \in \mathfrak{B}$ since $a_i \leq c$. So, $\mathfrak{B} \models \exists x \theta_2^i(x)$.

Hence, $\mathfrak{B} \models \theta$, as required. \(\square\)

**Claim 3.5.2.** $c \in \mathcal{K}_1^1(\mathfrak{B})$.

**Proof.** Note that $c = (\mu z)(\varphi((z)_0, (z)_1, b))$ in $\mathfrak{B}$, with $b \in I_1(\mathfrak{A})$. But it follows from $\mathfrak{B} \preceq_0 \mathfrak{A}$ and $\mathcal{K}_1^1(\mathfrak{A}) \subseteq \mathfrak{B}$ that $\mathcal{K}_1^1(\mathfrak{B}) = \mathcal{K}_1^1(\mathfrak{A})$. So, $b \in I_1(\mathfrak{B})$ and hence $c \in \mathcal{K}_1^1(\mathfrak{B})$. \(\square\)

**Claim 3.5.3.** $\mathfrak{B} \not\models T + f_{\Pi_2}(T + \text{Rfn}_{\Sigma_2}(T))$.

**Proof.** Pick $c \in \mathcal{K}_1^1(\mathfrak{A})$ nonstandard. Since $f(x) \geq 2^x \geq x$, $EA$ proves that $y = f^e(z)$ defines a nondecreasing function in the variable $x$. So, $f^e(c)$ does not exist in $\mathfrak{B}$. Consequently, $T + f_{\Pi_2}(T + \text{Rfn}_{\Sigma_2}(T))$ fails in $\mathfrak{B}$ by Proposition 3.4. \(\square\)

This completes the proof of the theorem. \(\square\)

For r.e. theories of arithmetical complexity $\mathcal{B}(\Sigma_1)$, one can obtain a particularly clean formulation of the previous theorem.
Theorem 3.6. Suppose $T \subseteq B(\Sigma_1)$. Then, $Th_{\Pi_2}(\text{Rfn}_{\Sigma_2}(T))$ is not contained in any consistent, r.e. extension of $T$ by $B(\Sigma_1)$–sentences.

Proof. Put $T = EA + \Gamma_1$, where $\Gamma_1$ is an r.e. set of $B(\Sigma_1)$–sentences. Consider $\Gamma_2 \subseteq B(\Sigma_1)$ r.e. and consistent with $EA + \Gamma_1$. By Theorem 3.5 $EA + \Gamma_1 + \Gamma_2 \not\vdash Th_{\Pi_2}(\text{Rfn}_{\Sigma_2}(EA))$. Consequently, $T + \Gamma_2 \not\vdash Th_{\Pi_2}(\text{Rfn}_{\Sigma_2}(T))$. □

Finally, we can now derive the answers to Question 1 and 2 as direct corollaries of our unboundedness theorem.

Corollary 3.7 (Answer to Question 1). Suppose $T_1$ is a finite $\Pi_2$–extension of $EA$ closed under $\Sigma_1$–CR, $T_2$ is an r.e. set of $B(\Sigma_1)$–sentences and $T = T_1 + T_2$. Then $T + \text{Rfn}_{\Sigma_2}(T)$ is not $\Pi_2$–conservative over $T + \text{Rfn}_{\Sigma_2}(T)$ provided $T + \text{Rfn}_{\Sigma_2}(T)$ is consistent.

Proof. It follows from Theorem 3.5 for $T_1$, because $T + \text{Rfn}_{\Sigma_2}(T)$ is a consistent, r.e. extension of $T_1$ by $B(\Sigma_1)$–sentences. □

Corollary 3.8 (Answer to Question 2). Suppose $S$ is a consistent, r.e. extension of $EA + \text{Rfn}_{\Sigma_2}(EA)$. Then $Th_{\Pi_2}(Th_{\Sigma_2}(S))$ is strictly stronger than $Th_{B(\Sigma_1)}(S)$.

Proof. By Theorem 3.6 for $T = EA$, $EA + Th_{B(\Sigma_1)}(S) \not\vdash Th_{\Pi_2}(\text{Rfn}_{\Sigma_2}(EA))$ and so $Th_{B(\Sigma_1)}(S) \not\vdash Th_{\Pi_2}(Th_{\Sigma_2}(S))$. □

4. Some Results à la Kreisel–Lévy

In [14] G. Kreisel and A. Lévy showed that $PA$ is equivalent to the full uniform reflection principle over primitive recursive arithmetic. D. Leivant and H. Ono ([15], [16]) sharpened Kreisel–Lévy’ result by establishing an exact correspondence between the hierarchy of $\Sigma_n$–induction and the hierarchy of partial uniform reflection principles, i.e., $EA + \text{RFN}_{\Sigma_{n+1}}(EA) \equiv I\Sigma_n$ for $n > 0$. Beklemishev ([2], [5]) extended this correspondence by showing that theories axiomatized by various forms of induction rules as well as parameter free induction schemes can also be characterized in terms of reflection principles. In particular, he proved that $\Pi_1^1$ is equivalent to $\text{Rfn}_{\Sigma_2}(EA)$, over the base theory $EA^+$ (which is $EA$ augmented with a $\Pi_2$–axiom declaring the totality of the superexponentiation function).

Theorem 4.1 (Beklemishev, [5]). Suppose $T$ is a finite $\Pi_2$–extension of $EA$.

(1) Over $EA$, it holds that $T + \text{Rfn}_{\Sigma_2}(T) \vdash T + \Pi_1^1$.

(2) Over $EA^+$, it holds that $T + \text{Rfn}_{\Sigma_2}(T) \equiv T + \Pi_1^1$.
In this section we show that the $\Gamma$–consequences of $T + \text{Rfn}_{\Sigma_2}(T)$, for $\Gamma = \Pi_2, B(\Sigma_1)$, can be characterized by local versions of the $\Sigma_1$–induction rule.

- The $\Sigma_1$–induction rule up to $K_1$ with parameters in $I_1^1$, denoted $(\Sigma_1, K_1, I_1^1)$–IR, is the inference rule given by:
  \[
  \frac{\varphi(0, v) \land \forall x \left( \varphi(x, v) \rightarrow \varphi(x + 1, v) \right)}{\forall v \in I_1^1 \forall x \in K_1 \varphi(x, v)},
  \]
  where $\varphi(x, v) \in \Sigma_1$.

- The parameter–free $\Sigma_1$–induction rule up to $K_1$, denoted $(\Sigma_1^-, K_1)$–IR, is the inference rule given by:
  \[
  \frac{\varphi(0) \land \forall x \left( \varphi(x) \rightarrow \varphi(x + 1) \right)}{\forall x \in K_1 \varphi(x)},
  \]
  where $\varphi(x) \in \Sigma_1^-$.

To prove our characterization results, we use a model–theoretic method inspired in J. Avigad’s [1] who, in turn, builds on some ideas of A. Visser (unpublished) and D. Zambella ([18]) in the context of Bounded Arithmetic. In [1] Avigad introduces the general notion of a Herbrand saturated model and shows that it provides us with a recipe for obtaining $\forall\exists$–conservation over universal theories. Here, we consider a hierarchical version of this notion that yields an unified method for proving $\Pi_n+1$–conservation over $\Pi_{n+2}$–theories. Recall that we write $\mathfrak{A} \prec_n \mathfrak{B}$ to mean that $\mathfrak{A}$ is a $\Sigma_n$–elementary substructure of $\mathfrak{B}$, i.e. for all $\varphi(x) \in \Sigma_n$ and $a \in \mathfrak{A}$, $\mathfrak{A} \models \varphi(a)$.

**Definition 4.2.** We say that $\mathfrak{A}$ is $\Sigma_{n+1}$–closed with respect to (w.r.t.) a theory $T$ if $\mathfrak{A} \models T$ and, for every $\mathfrak{B} \models T$,

$$\mathfrak{A} \prec_n \mathfrak{B} \Rightarrow \mathfrak{A} \prec_{n+1} \mathfrak{B}.$$  

First of all, note that, for all $n \geq 0$, $\Sigma_{n+1}$–closed models do exist. Indeed, a union of chain argument gives us that if $T \subseteq \Pi_{n+2}$ then every model of $T$ can be $\Sigma_n$–elementary extended to a $\Sigma_{n+1}$–closed model w.r.t. $T$. Next lemma is an analog of theorem 3.4 of [1].

**Lemma 4.3.** Suppose $T_2 \subseteq \Pi_{n+2}$. If every $\Sigma_{n+1}$–closed model w.r.t. $T_2$ is a model of $T_1$, then $T_1$ is $\Pi_{n+1}$–conservative over $T_2$.

Now we are in a position to prove our theorem. As usual, given an inference rule $R$ and a theory $T$, $T + R$ denotes the closure of $T$ under $R$ and first order logic; while $[T, R]$ denotes the closure of $T$ under non–nested applications of $R$ and first order logic. A rule $R_1$ is reducible to $R_2$ if $[T, R_1] \subseteq [T, R_2]$ for every theory $T$ extending $\text{EA}$; two rules $R_1$ and $R_2$ are congruent, written $R_1 \cong R_2$, if they are mutually reducible to each other.
Theorem 4.4. Let $T$ be a finite $\Pi_2$-extension of $EA$ closed under $\Sigma_1$–CR. Over $EA^+$, it holds that

1. $Th_{\Pi_2}(T + Rfn_{\Sigma_2}(T)) \equiv T + (\Sigma_1, K_1, T_1^1)–IR \equiv [T, (\Sigma_1, K_1, T_1^1)–IR]$.
2. $T + Th_{\Pi_2}(T + Rfn_{\Sigma_2}(T)) \equiv T + (\Sigma_1, K_1)–IR \equiv [T, (\Sigma_1, K_1)–IR]$.

Proof. (1): For notational simplicity, we write $T$ hence it is sufficient to show that $S$ Note that $D$ and hence $\phi$ $A$ is $\Sigma_1$–$IR$, and hence $\phi$ shows that every $\Sigma_1$–sentences. Thus, by theorem 3.2 of $[4]$ $S + B\Sigma_1$ is $\Pi_2$–conservative over $S$ and hence it is sufficient to show that $S + B\Sigma_1$ is closed under $(\Sigma_1, K_1, T_1^1)$–IR. To this end, assume

$S + B\Sigma_1 \vdash \phi(0, v) \land \forall x (\phi(x, v) \rightarrow \phi(x + 1, v))$,

with $\phi \in \Sigma_1$. Consider $\phi'(x, w) \equiv \forall v \leq w \phi(x, v)$. Clearly, $S + B\Sigma_1$ also proves the antecedent of the induction axiom for $\phi'$, and $\phi'$ is equivalent to a $\Sigma_1$–formula in $B\Sigma_1$. Since $S \models I(\Sigma_1, K_1, K_1^1)$ by Proposition 2.3,

$S + B\Sigma_1 \vdash \forall w \in K_1^1 \forall x \in K_1 \phi'(x, w)$

and so $\forall v \in T_1 \forall x \in K_1 \phi(x, v)$ is also provable in $S + B\Sigma_1$, as required.

$(R_\omega \vdash Th_{\Pi_2}(S))$: By Proposition 2.1, Theorem 4.1 and Lemma 4.3 it suffices to show that every $\Sigma_2$–closed model w.r.t. $R_\omega$ satisfies $I(\Sigma_1, K_1)$. To this end, assume $A$ is $\Sigma_2$–closed w.r.t. $R_\omega$, $a \in K_1(A)$ and $A \models \phi(0) \land \forall x (\phi(x) \rightarrow \phi(x + 1))$, with $\phi \in \Sigma_1$. We need to show $A \models \phi(a)$. We distinguish two cases.

Case 1: $T_1^1(A) \neq A$. Pick $b \in A$ such that $K_1^1(A) < b$. Put $\phi(x) \equiv \exists y \phi_0(x, y)$, with $\phi_0 \in \Delta_0$. It follows from $K_1^1(A) \prec_1 A$ that

$A \models \forall x \leq a (\exists y \phi_0(x, y) \leftrightarrow \exists y \leq b \phi_0(x, y))$.

Thus, $\phi(a)$ follows by applying $\Delta_0$–induction to $x \leq a \rightarrow \exists y \leq b \phi_0(x, y)$.

Case 2: $T_1^1(A) = A$. It follows from the fact that $A$ is $\Sigma_2$–closed w.r.t. $R_\omega$ that $R_\omega + D\Pi_1(A) \vdash Th_{\Pi_2}(A)$, where $D\Gamma(A)$ denotes the $\Gamma$–diagram of $A$, i.e. the set of all $\Gamma$–formulas (possibly with parameters from $A$) which are true in $A$. Thus, by compactness there is $\delta(b) \in D\Pi_1(A)$ satisfying that

$R_\omega + \delta(b) \models \phi(0) \land \forall x (\phi(x) \rightarrow \phi(x + 1))$,

and hence

$R_\omega \vdash \forall v (\delta(v) \rightarrow [\phi(0) \land \forall x (\phi(x) \rightarrow \phi(x + 1))])$.

Put $\phi'(x, v) \equiv \delta(v) \rightarrow \phi(x)$. Obviously, $\phi' \in \Sigma_1$ and $R_\omega$ proves the antecedent of the induction axiom for $\phi'$.

Since $R_\omega$ is closed under $(\Sigma_1, K_1, T_1^1)$–IR and trivially $b \in T_1^1(A)$, we have $A \models \phi'(a, b)$ and so $A \models \phi(a)$.
(R₁ ⊨ Rω): We need to show that over T, nested applications of (Σ₁, K₁, I₁₁)–IR, Rω, can be reduced to non–nested applications of the rule, R₁. To see this, assume $\mathfrak{A} \models R₁$. We distinguish two cases.

Case 1: $I₁₁(\mathfrak{A}) \neq \mathfrak{A}$. Reasoning as in case 1 above, we get that $\mathfrak{A}$ satisfies $T + \Pi₁$ and hence also $S = T + \text{Rfl}_{Σ₁}(T)$. Thus, the result follows because we have already shown that $Rω \subseteq S$.

Case 2: $I₁₁(\mathfrak{A}) = \mathfrak{A}$. Then $\mathfrak{A}$ also satisfies $[T, (Σ₁, K₁)–IR]$, where $(Σ₁, K₁)–IR$ denotes the stronger inference rule of parametric $Σ₁$–induction up to $K₁$, given by:

$$\varphi(0, v) \land \forall x (\varphi(x, v) \rightarrow \varphi(x + 1, v)) \quad \forall x ∈ K₁ \forall v \varphi(x, v),$$

where $\varphi(x, v)$ is in $Σ₁$. Since $Rω$ is clearly contained in $T + (Σ₁, K₁)–IR$, it suffices to show that $T + (Σ₁, K₁)–IR$ collapses to non–nested applications of the rule $[T, (Σ₁, K₁)–IR]$. To this end, put $T \equiv EA + \forall x ∃y (y = f(x))$, where $f$ is $EA$–honest. The key idea is that $[T, (Σ₁, K₁)–IR]$ is equivalent to a certain form of restricted iteration, namely, the x-th iterate of $f$ is total whenever $x$ is $Σ₁$–definable.

For the sake of readability, we shall write $f \downarrow$ for the $\Pi₂$–axiom $\forall x ∃y (y = f(x))$ and $f^x \downarrow$ for $\forall z ∃y (y = f^x(z))$.

Claim 4.4.1. Suppose $y = f(x)$ is $EA$–honest. Then

$$[EA + f \downarrow, (Σ₁, K₁)–IR] \equiv EA + \forall x ∈ K₁ f^x \downarrow.$$

The proof of the claim is a rather standard argument using Parikh’s theorem and we omit it (we refer the reader to lemma 4.8 of [9] for a detailed account of the proof). Equipped with this result, a proof of the collapse of $(Σ₁, K₁)–IR$ naturally arises: nested applications of the rule correspond to nested iteration of $f$ up to $Σ₁$–definable elements; but $(f^{x₁})^{x₂}(z) = f^{x₁ \cdot x₂}(z)$ and the product of two $Σ₁$–definable elements is $Σ₁$–definable too. More formally, assume that $\mathfrak{B}$ is a model of $[T, (Σ₁, K₁)–IR]$ and $\varphi(x, 0) \land \forall x (\varphi(x, v) \rightarrow \varphi(x + 1, v))$ is provable in $[T, (Σ₁, K₁)–IR]$, with $\varphi ∈ Σ₁$.

We must show $\mathfrak{B} \models \forall x ∈ K₁ \forall v \varphi(x, v)$. For a formula $δ(x)$, recall that $Def_δ(x)$ denotes $δ(x) \land \forall x, x′ (δ(x) ∧ δ(x′) → x = x′)$. By compactness, it follows from claim 4.4.1 that there are $δ₁(x), . . . , δ_k(x) ∈ Σ₁$ satisfying that

$$EA + \bigwedge_{i=1}^{k} \forall x (Def_δ_i(x) → f^x \downarrow) \vdash \varphi(0, 0) \land \varphi(x, v) \rightarrow \varphi(x + 1, v)).$$

Now we divide the set of formulas $δ₁(x), . . . , δ_k(x)$ into two subsets: those for which $∃x δ_i(x)$ holds in $\mathfrak{B}$, and those for which it does not. Without loss of generality, we may assume that the first subset consists of $δ₁(x), . . . , δ_m(x)$, with $m ≤ k$. For each $i ≤ m$ put $δ_i(x) = ∃y δ_i′(x, y)$, with $δ_i′ ∈ Δ₀$. Let $b_i$ denote the element defined by $δ_i(x)$ and let $b_i′$ denote the least element satisfying $∃x, y ≤ z (z = (x, y) ∧ δ_i′(x, y))$. It
is clear that each $b'_i$ is $\Delta_0$-definable and $b_i \leq b'_i$. Let $\delta(x)$ be a $\Delta_0$-formula defining $b = b'_1 + \cdots + b'_m$ and put

$$y = g(x) \equiv \exists u \leq y (\delta(u) \land f^u(x) = y).$$

Finally, let $\Theta$ denote $\bigwedge_{i=1}^{n} \forall x - \delta_i(x)$. Clearly, $\mathfrak{B} \models \Theta$ and $EA + \Theta + g \downarrow$ implies $\bigwedge_{i=1}^{n} \forall x (Def_{A}(x) \rightarrow f^x(x))$. Hence, we have

$$[EA + \Theta + g \downarrow, (\Sigma_1, \mathcal{K}_1)-IR] \vdash \forall x \in \mathcal{K}_1 \forall v \varphi(x, v);$$

so, since $\Theta \subseteq \Pi_1$,

$$[EA + g \downarrow, (\Sigma_1, \mathcal{K}_1)-IR] + \Theta \vdash \forall x \in \mathcal{K}_1 \forall v \varphi(x, v);$$

and therefore, since $y = g(x)$ is $EA$-honest,

$$EA + \Theta + \forall x \in \mathcal{K}_1 g^x \downarrow \vdash \forall x \in \mathcal{K}_1 \forall v \varphi(x, v)$$

by Claim 4.4.1. But $\mathfrak{B} \models \forall x \in \mathcal{K}_1 g^x \downarrow$ since, provably in $EA + g \downarrow$, $y = g^x(z)$ is equivalent to $\exists u \leq y (\delta(u) \land y = f^x(u(z)))$.

(2): For simplicity, we write $R^-_\omega$ for $T + (\Sigma_1, \mathcal{K}_1)$-IR, and $R^-_1$ for $[T, (\Sigma_1, \mathcal{K}_1)-IR]$. $(T + Th_{B(\Sigma_1)}(S) \vdash R^-_\omega)$: Obviously, $R^-_\omega \subseteq R^-_\omega$ and so $Th_{\Pi_2}(S) \vdash R^-_\omega$ by part (1). Note that if $\varphi(x) \in \Sigma_1$ then $\forall x \in \mathcal{K}_1 \varphi(x)$ is equivalent in $EA$ to

$$\{ \exists x Def_{b}(x) \rightarrow \exists x (Def_{b}(x) \land \varphi(x)) : \delta(x) \in \Sigma_1 \},$$

which can be rewritten as a set of $B(\Sigma_1)$-sentences. Hence, $R^-_\omega \subseteq T + Th_{B(\Sigma_1)}(S)$, as required.

$(R^-_\omega \vdash R^-_1)$: Immediate.

$(R^-_1 \vdash Th_{B(\Sigma_1)}(S))$: Suppose $\mathfrak{A} \models R^-_1$. It follows from $\mathcal{K}_1(\mathfrak{A}) \prec_1 \mathfrak{A}$ that $\mathcal{K}_1(\mathfrak{A}) \models [T, \Sigma_1-IR]$. But it is well-known that $\Sigma_1$-IR and $\Sigma_1^-\text{-IR}$ are congruent rules (see, e.g., lemma 2.1 in [5]) and hence $\mathcal{K}_1(\mathfrak{A}) \models [T, \Sigma_1-IR]$ too. Since $\mathcal{K}_1(\mathfrak{A}) \models T_{\Pi_2}(S)$ by part (1) and so $\mathfrak{A} \models Th_{B(\Sigma_1)}(S)$.

Remark 4.5. Note that in the proof of Theorem 4.4 $EA^+$ is only needed to show that the induction rule theories imply the reflection principle theories, as we make use of part (2) of Theorem 4.1. The rest of the implications in the statement of the theorem are also true over the weaker theory $EA$.

Interestingly, $(\Sigma_1^-, \mathcal{K}_1)$-IR can be reformulated in terms of a more “classic” inference rule. Let $\Gamma-\text{IR}_0$ denote the variant of the usual induction rule given by:

$$\frac{\forall x (\varphi(x) \rightarrow \varphi(x + 1))}{\varphi(0) \rightarrow \forall x \varphi(x)},$$

where $\varphi(x) \in \Gamma$. In his detailed analysis of induction rules in arithmetic [2] Beklemishev proved that, if parameters are allowed, IR$_0$ is congruent with the usual
Let $\Sigma_n^\perp -IR_0 \equiv \Pi_n^\perp -IR_0 \equiv \Sigma_n^\perp -IR$. The parameter free version of $IR_0$ was not considered there. However, it turns out that, whereas $\Sigma_n^\perp -IR_0$ is easily seen to be congruent with $\Sigma_n^\perp -IR_0$, $\Pi_n^\perp -IR_0$ is no longer reducible to its parameter free counterpart. In fact, next result shows that $\Pi_1^\perp -IR_0$ coincides, precisely, with our $(\Sigma_1^\perp, K_1^\perp)$–IR.

**Proposition 4.6.** $\Pi_1^\perp -IR_0$ and $(\Sigma_1^\perp, K_1^\perp)$–IR are congruent rules.

**Proof.** $(\Rightarrow)$: Suppose $\mathfrak{A} \models [T, \Pi_1^\perp -IR_0], T \vdash \varphi(0) \land \forall x \ ( \varphi(x) \rightarrow \varphi(x + 1) )$, with $\varphi(x) \in \Sigma_1^\perp$, and $a \in K_1(\mathfrak{A})$. Towards a contradiction, assume $\mathfrak{A} \vdash \neg \varphi(a)$. Let $\delta(z) \in \Sigma_1$–formula defining $a$ and put $\theta(x) \equiv \forall z \ ( \delta(z) \rightarrow \neg \varphi(z - x) )$. It is clear that $\theta(x) \in \Pi_1^\perp$ and $T$ proves $\theta(x) \rightarrow \theta(x + 1)$. By applying $\Pi_1^\perp -IR_0$, we get $\mathfrak{A} \models \theta(0) \rightarrow \forall x \ \theta(x)$. But $\mathfrak{A} \models \theta(0)$ since $\varphi(a)$ fails in $\mathfrak{A}$. Hence, $\mathfrak{A} \models \theta(a)$ and so $\mathfrak{A} \models \neg \varphi(0)$, which is a contradiction.

$(\Leftarrow)$: Suppose $\mathfrak{A} \models [T, (\Sigma_1^\perp, K_1^\perp)$–IR]. It follows from $K_1(\mathfrak{A}) \prec_1 \mathfrak{A}$ that $K_1(\mathfrak{A}) \models [Th_{\Pi_2}(T), \Sigma_1^\perp -$IR]. So, $K_1(\mathfrak{A}) \models [Th_{\Pi_2}(T), \Pi_1^\perp -IR_0]$ since $\Sigma_1^\perp -$IR and $\Pi_1^\perp -$IR$_0$ are congruent rules. As a result, $\mathfrak{A} \models [T, \Pi_1^\perp -$IR$_0]$, for the latter theory is an extension of $T$ by $B(\Sigma_1)$–sentences.

As an application, we are able to derive a Kreisel and Lévy–like result for local $\Sigma_1$–reflection, thus filling a gap in our understanding of the equivalence between local reflection and induction.

**Theorem 4.7.** Let $T$ be a finite $\Pi_2$–extension of $EA$. Then,

\[ T + \text{Rfn}_{\Sigma_1}(T) \equiv T + \Pi_1^\perp -IR_0 \equiv [T, \Pi_1^\perp -$IR$_0]. \]

**Proof.** We write $S$ for $T + \text{Rfn}_{\Sigma_1}(T)$, $R^-_\omega$ for $T + \Pi_1^\perp -$IR$_0$, and $R^-_1$ for $[T, \Pi_1^\perp -$IR$_0]$.

$(S \vdash R^-_\omega)$: By Theorem 4.1, $T + \text{Rfn}_{\Sigma_1}(T) \vdash T + \Pi_1^- \vdash R^-_\omega$. Since the latter theory is an extension of $T$ by $B(\Sigma_1)$–sentences, the result follows from $B(\Sigma_1)$–conservativity of $T + \text{Rfn}_{\Sigma_1}(T)$ over $S$.

$(R^-_\omega \vdash R^-_1)$: Immediate.

$(R^-_1 \vdash S)$: Suppose $\mathfrak{A} \models R^-_1$. It follows from Proposition 4.6 and $K_1(\mathfrak{A}) \prec_1 \mathfrak{A}$ that $K_1(\mathfrak{A}) \models [T, \Sigma_1 -$IR]. By theorem 2 of [2], $[T, \Sigma_1 -$IR] implies $T + \text{Rfn}_{\Sigma_1}(T)$. Hence, $\mathfrak{A} \models T + Th_{\Sigma_2}(T + \text{Rfn}_{\Sigma_1}(T))$ and so $\mathfrak{A} \models S$. \hfill $\Box$

**Corollary 4.8.** $EA + \text{Rfn}_{\Sigma_1}(EA) \equiv [EA, \Pi_1^\perp -$IR$_0]$. 

Combining Theorem 4.7 and the closed model method for proving conservativity, we can also obtain a proof that the $\Pi_1$–consequences of local reflection correspond to the induction rule for $\Pi_1$–formulas $\Pi_1$–IR, a result that, in fact, is a corollary of theorem 1 in [2].
Theorem 4.9. Let $T$ be a finite $\Pi_2$-extension of $EA$. Then,
\[ T + Th_{\Pi_1}(T + Rfn(T)) \equiv T + \Pi_1-IR. \]

Proof. We write $S$ for $T + Rfn(T)$ and $R_\omega$ for $T + \Pi_1-IR$.

($T + Th_{\Pi_1}(S) \vdash R_\omega$): By Theorem 4.1, $T + Rfn(T) \vdash T + \Pi_1^- \vdash R_\omega$. So, $T + Th_{\Pi_1}(S)$ implies $R_\omega$ because $R_\omega$ is an extension of $T$ by $\Pi_1-$sentences.

($R_\omega \vdash Th_{\Pi_1}(S)$): By Theorem 1.1, $S$ and $T + Rfn_{\Sigma_1}(T)$ prove the same $\Pi_1-$sentences. So, it follows from Lemma 4.3 and Theorem 4.7 that every $\Sigma_1$-closed model w.r.t. $R_\omega$ satisfies $\exists \delta(a)$ in the $\Pi_0$-diagram of $\mathfrak{A}$ satisfying that $R_\omega + \delta(a) \vdash \varphi(0)$.

Put $\varphi'(x) \equiv \exists z \delta(z) \rightarrow \varphi(x)$. Clearly, $\varphi'(x) \in \Pi_1^-$ and $R_\omega$ proves the antecedent of the induction axiom for $\varphi'$. Hence, $\mathfrak{A} \models \forall x \varphi'(x)$ and so $\mathfrak{A} \models \forall x \varphi(x)$, as required.

According to a theorem of S.V. Goryachev [10] (see also proposition 2.33 of [7]) $T + Rfn(T)$ is a $\Pi_1$-conservative extension of the so-called $\omega$ times iterated consistency assertion for $T$

\[ T + Con(T) + Con(T + Con(T)) + \ldots \]

As a result, Theorem 4.9 also gives us that

Corollary 4.10. Let $T$ be a finite $\Pi_2$-extension of $EA$. Then,
\[ T + \Pi_1-IR \equiv T + Con(T) + Con(T + Con(T)) + \ldots \]

The previous characterization of $\Pi_1-IR$ was obtained in [2] for finite $\Pi_2$-extensions of $EA^+$ (see corollary 3.5 there). $EA^+$ is needed because the proof given in [2] uses formalized Cut-elimination Theorem. In fact, the author remarked that, for obtaining a characterization modulo congruence in a theory weaker than $EA^+$ one should replace $Con(T)$ with a suitable weaker notion of consistency (e.g. $Con^{cf}(T)$ expressing the consistency of $T$ in the sense of the cut-free provability). Nonetheless, taking $T = EA$ in Corollary 4.10 we obtain that the result remains true modulo $EA$ even for the standard notion of consistency.

Corollary 4.11.
\[ EA + \Pi_1-IR \equiv EA + Con(EA) + Con(EA + Con(EA)) + \ldots \]
5. The case \( n > 0 \)

It is natural to ask ourselves how our results generalize to \( \text{Rfn}_{\Sigma_{n+2}}(T) \) and \( \Pi_{n+1}^- \) for an arbitrary \( n > 0 \). First of all, it should be noticed that, in order to characterize \( \Pi_{n+1}^- \) in terms of reflection principles, one needs to consider relativized local reflection. For each \( n > 0 \), the relativized local reflection principle for \( T \) is the scheme given by

\[
\text{Rfn}^n(T) : [n]_T(⌜\varphi⌝) \rightarrow \varphi,
\]

where \( \varphi \) ranges over all sentences in \( \Gamma \) and \( [n]_T(x) \) denotes a \( \Sigma_{n+1} \)–formula expressing “\( x \) is provable from \( T + Th_{\Pi_n}(N) \).” That is to say, \( [n]_T(x) \equiv \upsilon \), where

\[
Ax_{\upsilon}(x) = (Ax_T(x) \lor \text{True}_{\Pi_n}(x))
\]

and \( \text{True}_{\Pi_n}(x) \) is a truth–definition for \( \Pi_n \)–sentences in \( EA \) (see section 2.3 of [7] for details). It is a theorem of Beklemishev that relativized local reflection principles capture the hierarchy of parameter free fragments of \( PA \). In fact, in [5] it is shown that over \( EA \), \( \Pi_{n+1}^n \equiv \text{Rfn}^n_{\Sigma_{n+2}}(EA) \) and \( I_{\Sigma_{n+1}}^n \equiv \text{Rfn}^n_{\Sigma_{n+1}}(EA) \) for each \( n > 0 \).

In addition, a relativized version of the conservation Theorem 1.1 holds, i.e. \( T + \text{Rfn}^n_{\Sigma_{n+2}}(T) \) is \( B(\Sigma_{n+1}) \)–conservative over \( T + \text{Rfn}^n_{\Sigma_{n+1}}(T) \). Thus, a relativized version of Question 1 is in order.

**Question 3** \((n > 0)\). Is \( T + \text{Rfn}^n_{\Sigma_{n+2}}(T) \) conservative over \( T + \text{Rfn}^n_{\Sigma_{n+1}}(T) \) with respect to \( \Pi_{n+2} \)–sentences?

It turns out that the ideas in the previous sections apply equally well to the case \( n > 0 \) for theories \( T \) extending the \( \Sigma_n \)–induction scheme \( I_{\Sigma_n} \) and closed under \( \Sigma_{n+1} \)–collection rule. Such a theory \( T \) can be reformulated as \( I_{\Sigma_n} + “f \text{ is total}”, \) where \( f \) is a nondecreasing function with a \( \Pi_n \)–graph. Thus, considering the \( \Pi_{n+2} \) separation property

\[
\forall x \in I_{\Pi_{n+1}} \forall z \in \mathcal{K}_{n+1} \exists y \varphi(x, y, z) \exists s \forall x \leq z \varphi(x, (s)x, v),
\]

allows us to obtain an analog of the unboundedness Theorem 3.5 for the class \( Th_{\Pi_{n+2}}(\text{Rfn}^n_{\Sigma_{n+2}}(T)) \) and, in turn, a negative answer to Question 3.

It will be useful, however, to obtain an answer to Question 3 for theories \( T \) extending \( EA \) rather than extending \( I_{\Sigma_n} \). To this end we have to use a different separation property. Again, the key point is considering arithmetic theories up to definable elements, namely, the following local variant of the *finite axiom of choice* for \( \Pi_{n-1} \)–formulas.

- The scheme of finite axiom of choice for \( \Pi_{n-1} \)–formulas up to \( \mathcal{K}_1 \) with parameters in \( \mathcal{K}_{n+1} \), denoted \( F\text{AC}(\Pi_{n-1}, \mathcal{K}_1, \mathcal{K}_{n+1}) \), is \( EA \) plus

\[
\forall v \in \mathcal{K}_{n+1} \forall x \in \mathcal{K}_1 (\forall x \leq z \exists y \varphi(x, y, v) \rightarrow \exists s \forall x \leq z \varphi(x, (s)x, v)),
\]

where \( \varphi \) is in \( \Pi_{n-1} \).
It is easy to check that $FAC(\Pi_{n-1}, K_1, \mathcal{K}^1_{n+1})$ can be reexpressed as a set of $\Pi_{n+2}$-sentences which are provable from $\Pi_{n+1}^-$ by Proposition 2.1 (and hence also from $T + \text{Rfn}^{n}_{\Sigma_{2n+2}}(T)$). By using this separation property, we obtain the desired unboundedness theorem for theories $T$ extending $EA$.

**Theorem 5.1 (n > 0, Unboundedness).** Suppose $T$ is a $B(\Sigma_{n+1})$-extension of $EA$. Then, $\text{Th}_{\Pi_{n+2}}(\text{Rfn}^{n}_{\Sigma_{n+2}}(T))$ is not contained in any consistent, r.e. extension of $T$ by $B(\Sigma_{n+1})$-sentences.

**Proof.** is similar to that of Theorem 3.5 so we skip some details. Towards a contradiction, assume that there is an r.e. set of $B(\Sigma_{n+1})$-sentences $\Gamma$ such that $T + \Gamma$ is consistent and contains $\text{Th}_{\Pi_{n+2}}(\text{Rfn}^{n}_{\Sigma_{n+2}}(T))$. In particular, $T + \Gamma + \Sigma_n^-$ is consistent and then so is $T + \Gamma + I \Sigma_n$, for $I \Sigma_n$ is $\Sigma_{n+2}$-conservative over $\Sigma_n^-$ by theorem 2.1 of [13]. Let $\mathfrak{A}$ be a model of $T + \Gamma + I \Sigma_n$ with $K_1(\mathfrak{A}) \neq \omega$ and $\mathcal{I}_{n+1}(\mathfrak{A}) \neq \mathfrak{A}$. Since $\mathfrak{A} \models I \Sigma_n$, every proper initial $I \Sigma_n$-substructure of $\mathfrak{A}$ satisfies $B \Sigma_{n+2}$. Hence, reasoning as in the proof of Theorem 3.5, we get that there is $a \in K^1_{n+1}(\mathfrak{A}) - \mathcal{I}_{n+1}(\mathfrak{A})$. It follows by an overspill argument inside $K^1_{n+1}(\mathfrak{A})$ that there is a sequence coded by $b \leq a$ such that, for all $\theta(x,y) \in \Pi_n$,

$$K^1_{n+1}(\mathfrak{A}) \models \exists x, y \leq a \text{Sat}_n(\langle \theta \rangle, x, y) \rightarrow (b)_{\langle \theta \rangle} = (\mu t) (\text{Sat}_n(\langle \theta \rangle, (t)_0, (t)_1)),$$

where $\text{Sat}_n$ is a truth predicate for $\Pi_n$-formulas in $EA$ and $(t)_0, (t)_1$ denote the inverse of the Cantor pairing function. Note that $K^1_{n+1}(\mathfrak{A})$ is included in any substructure of $\mathfrak{A}$ containing $b$, for in models of $I \Sigma_n$ every $\Sigma_{n+1}$-definable element can be obtained as the projection of a $\Pi_n$-minimal one. Since $b \in K^1_{n+1}(\mathfrak{A})$, there are $b_1 \in K^1_{n+1}(\mathfrak{A})$ and $b_2 \in I_{n+1}(\mathfrak{A})$ satisfying that $b_1 = (\mu t) (\varphi((t)_0, (t)_1, b_2))$ and $b = (b_1)_0$ for some $\varphi \in \Pi_n$. Put $c = \langle b_1, b_2 \rangle$ and define $\mathfrak{B}$ to be $K_n(\mathfrak{A}, c)$, i.e. the submodel of all elements which are $\Sigma_n$-definable using the parameter $c$. Then, we have:

i) $\mathfrak{B} \models T + \Gamma$. It follows from $\mathfrak{B} \subseteq \mathfrak{A}$, $K^1_{n+1}(\mathfrak{A}) \subseteq \mathfrak{B}$ and $T + \Gamma \subseteq B(\Sigma_{n+1})$.

ii) $c \in K^1_{n+1}(\mathfrak{B})$. It follows from $\mathfrak{B} \subseteq \mathfrak{A}$ and $I_{n+1}(\mathfrak{A}) \cap \mathfrak{B} = I_{n+1}(\mathfrak{B})$.

iii) $\mathfrak{B} \not\models FAC(\Pi_{n-1}, K_1, \mathcal{K}^1_{n+1})$. We argue as in Paris–Kirby’ proof that the $\Sigma_n$-collection scheme fails in $K_n(\mathfrak{A})$ (see proposition 7 of [17]). Let $Min_{n-1}(z, y, v)$ formalize “$z$ is the least element satisfying the $\Pi_{n-1}$-formula $y$ with a parameter $v$.” If $\mathfrak{B}$ were to satisfy $FAC(\Pi_{n-1}, K_1, \mathcal{K}^1_{n+1})$ then, for any nonstandard $\Sigma_1$-definable element $d$, there would be an element $e$ such that

$$\mathfrak{B} \models \forall x \leq d \exists y < d \exists z < e (Min_{n-1}(z, y, c) \land x = (z)_0),$$

violating the pigeon–hole principle for $\Sigma_0(\Sigma_{n-1})$–functions in models of $I \Sigma_{n-1}$. It follows from i) and iii) that $T + \Gamma$ does not imply $FAC(\Pi_{n-1}, K_1, \mathcal{K}^1_{n+1})$, which contradicts our assumption that $T + \Gamma$ contains $\text{Th}_{\Pi_{n+2}}(\text{Rfn}^{n}_{\Sigma_{2n+2}}(T))$. \hfill \Box
As a consequence, we obtain the corresponding negative answer to Question 3.

**Corollary 5.2** \((n > 0)\). Suppose \(T\) is a \(B(\Sigma_{n+1})\)-extension of \(EA\). \(T + Rfn^n_{\Sigma_{n+2}}(T)\) is not \(\Pi_{n+2}\)-conservative over \(T + Rfn^n_{\Sigma_{n+1}}(T)\) provided that the latter theory is consistent.

Thus, known \(B(\Sigma_{n+1})\)-conservativity of \(T + Rfn^n_{\Sigma_{n+2}}(T)\) over \(T + Rfn^n_{\Sigma_{n+1}}(T)\) for \(n > 0\) is optimal with respect to arithmetic complexity. In addition, Theorem 5.1 yields an application to the study of parameter free induction. In fact, taking \(T = EA\), our proof of Theorem 5.1 gives us that

**Corollary 5.3** \((n > 0)\). \(\Pi_{n+1}^-\) is not \(\Pi_{n+2}\)-conservative over \(I\Sigma^-_n\).

**Remark 5.4.**

1. In [8] we gave another proof of Corollary 5.3 that uses a different separation property, namely, a formalized version of the model-theoretic property

\[
(A) \equiv \forall a \in K^1_{n+1} \ "K_n(\mathfrak{A}, a) \ is \ bounded \ above \ in \ \mathfrak{A}.”
\]

This approach has the small advantage that it allows one to prove that \(\Pi_{n+1}^-\) is not \(\Pi_{n+2}\)-conservative over \(I\Sigma^-_n + B\Sigma_n\) for each \(n > 0\) (see the proof of theorem 5.5 in [8] for details).

2. Note that both \(FAC(\Pi_{n-1}, K_1, K^1_{n+1})\) and property \((A)\) above are provable from \(I\Sigma_n\). This fact is not coincidental as we shall infer from the following conservation result.

Let \(Rfn^0(T)\) coincide, by definition, with its non-relativized analog \(Rfn^0(T)\) and let \(Rfn^n(T)\) denote the full relativized local reflection principle for \(T\).

**Theorem 5.5** \((n \geq 0)\).

1. \(T + Rfn^n(T)\) is \(\Pi_{n+2}\)-conservative over \(T + RFN_{\Sigma_{n+1}}(T)\).

2. \(T + \Pi_{n+1}^-\) is \(\Pi_{n+2}\)-conservative over \([T, \Sigma_{n+1}IR]\).

**Proof.** (1): Let \(\varphi\) be a \(\Pi_{n+2}\)-sentence.

\[
T + Rfn^n(T) \vdash \varphi \quad \Rightarrow \quad T + Rfn^o_{I_{n+2}}(T) \vdash \varphi \quad \Rightarrow \quad T + RFN_{\Pi_{n+2}}(T) \vdash \varphi,
\]

where the first implication follows by a relativized version of Theorem 1.1; and the second one is trivial for \(n = 0\) or easily follows by using formalized Deduction theorem for \(n > 0\). Note that over \(EA\), \(RFN_{\Pi_{n+2}}(T) \equiv RFN_{\Sigma_{n+1}}(T)\) and hence the result follows.
Let \( \varphi \) be a \( \Pi_{n+2}^- \)–sentence provable in \( T + \Pi_{n+1}^- \). Since \( \Pi_{n+1}^- \subseteq \Sigma_{n+2} \), by compactness there is a finite \( \Pi_{n+2}^- \)–axiomatized subtheory \( T_0 \subseteq T \) satisfying that \( T_0 + \Pi_{n+1}^- \vdash \varphi \). Then

\[
T_0 + \Pi_{n+1}^- \vdash \varphi \implies T_0 + Rfn_{n+2}^0(T_0) \vdash \varphi
\]

\[
T_0 + \Pi_{n+1}^- \vdash \varphi \implies T_0 + Rfn_{n}(T_0) \vdash \varphi
\]

\[
T_0 + \Pi_{n+1}^- \vdash \varphi \implies T_0 + Rfn_{\Sigma_{n+1}}(T_0) \vdash \varphi
\]

by part (1). But \( [T_0, \Sigma_{n+1}^-\text{IR}] \equiv T_0 + Rfn_{\Sigma_{n+1}}(T_0) \) by theorem 4 of [2] and the result follows. \( \square \)

By using a model–theoretic construction Kaye, Paris and Dimitracopoulos proved that, when formulated in the usual language of arithmetic, \( \Pi_1^I \) is \( \Pi_2^- \)–conservative over \( I\Delta_0 + \text{exp} \equiv [I\Delta_0, \Sigma_1^-\text{IR}] \) (see theorem 2.9 of [13]). Thus, Theorem 5.5 can be seen as a general reflection principle counterpart of Kaye–Paris–Dimitracopoulos’ result. Also, note that, while \( I\Sigma_n \) and \( I\Sigma_n^- \) cease to be deductively equivalent. Therefore, two natural generalizations of Kaye–Paris–Dimitracopoulos’ result to \( n > 0 \) are in order. Firstly, taking \( T = I\Sigma_n \) in Theorem 5.5, we get that \( I\Sigma_n + \Pi_{n+1}^- \) is \( \Pi_{n+2}^- \)–conservative over \( [I\Sigma_n, \Sigma_{n+1}^-\text{IR}] \) (indeed, this is the one that, in an implicit manner, was considered in [13]). Secondly, since \( [EA, \Sigma_{n+1}^-\text{IR}] \equiv EA + Rfn_{\Sigma_{n+1}}(EA) \equiv I\Sigma_n \), taking \( T = EA \) in Theorem 5.5 yields the following new conservation result for induction schemes.

**Corollary 5.6** \( (n > 0) \). \( \Pi_{n+1}^- \) is \( \Pi_{n+2}^- \)–conservative over \( I\Sigma_n \).

Corollary 5.6 explains why any \( \Pi_{n+2}^- \)–property separating \( \Pi_{n+1}^- \) and \( I\Sigma_n^- \) turns out to be provable from \( I\Sigma_n \) as well as, combined with Corollary 5.3, settles the question of the optimality of conservativity between \( \Pi_{n+1}^- \) and \( I\Sigma_n^- \) for \( n > 0 \). On the one hand, \( \Pi_{n+1}^- \) is \( B(\Sigma_{n+1}) \)–conservative over \( I\Sigma_n^- \) and this result is best possible; on the other hand, conservativity of \( \Pi_{n+1}^- \) over \( I\Sigma_n \) can be extended to \( \Pi_{n+2}^- \)–sentences and this is, again, best possible.

**References**


22


