Notes on local reflection principles

Lev Beklemishev
Steklov Mathematical Institute
Vavilova 42, Moscow 117966, RUSSIA
e-mail: lev@bekl.mian.su

April 27, 1995

Abstract

We study the hierarchy of reflection principles obtained by restricting the full local reflection schema to the classes of the arithmetical hierarchy. Optimal conservation results w.r.t. the arithmetical complexity for such principles are obtained.

Reflection principles, for an arithmetical theory T, are formal schemata expressing the soundness of T, that is, the statement that "every sentence provable in T is true". More precisely, if $\mathrm{Prov}_T(x)$ denotes the canonical Σ^0_1 provability predicate for T, then the *local* reflection principle for T is the schema

$$\mathrm{Prov}_T(\ulcorner A \urcorner) \to A, \quad A \text{ is a sentence,}$$

and *uniform* reflection principle is the schema

$$\forall x \, (\mathrm{Prov}_T(\ulcorner A(\dot{x}) \urcorner) \to A(x)), \quad A(x) \text{ is a formula.}$$

We denote local and uniform reflection principles respectively Rfn_T and RFN_T. Other natural forms of reflection turn out to be equivalent to one of these two (cf also [8]). *Partial* reflection principles are obtained from local and uniform schemata by imposing a restriction that the formula A may only range over a certain subclass Γ of the class of T-sentences (formulas). Such schemata will be denoted $\mathrm{Rfn}_T(\Gamma)$ and $\mathrm{RFN}_T(\Gamma)$, respectively, and for Γ one usually takes one of the classes Σ^0_n or Π^0_n of the arithmetical hierarchy. $\mathcal{B}(\Sigma^0_n)$ denotes the class of all boolean combinations of Σ^0_n sentences.

In this note we consider some basic questions concerning the hierarchy of partial local reflection principles: the collapse of this hierarchy, finite axiomatizability of the theories of the hierarchy, etc. We also obtain optimal conservation results for partial local reflection principles. The corresponding questions for uniform reflection principles are well-known and easy, but are resolved in a rather

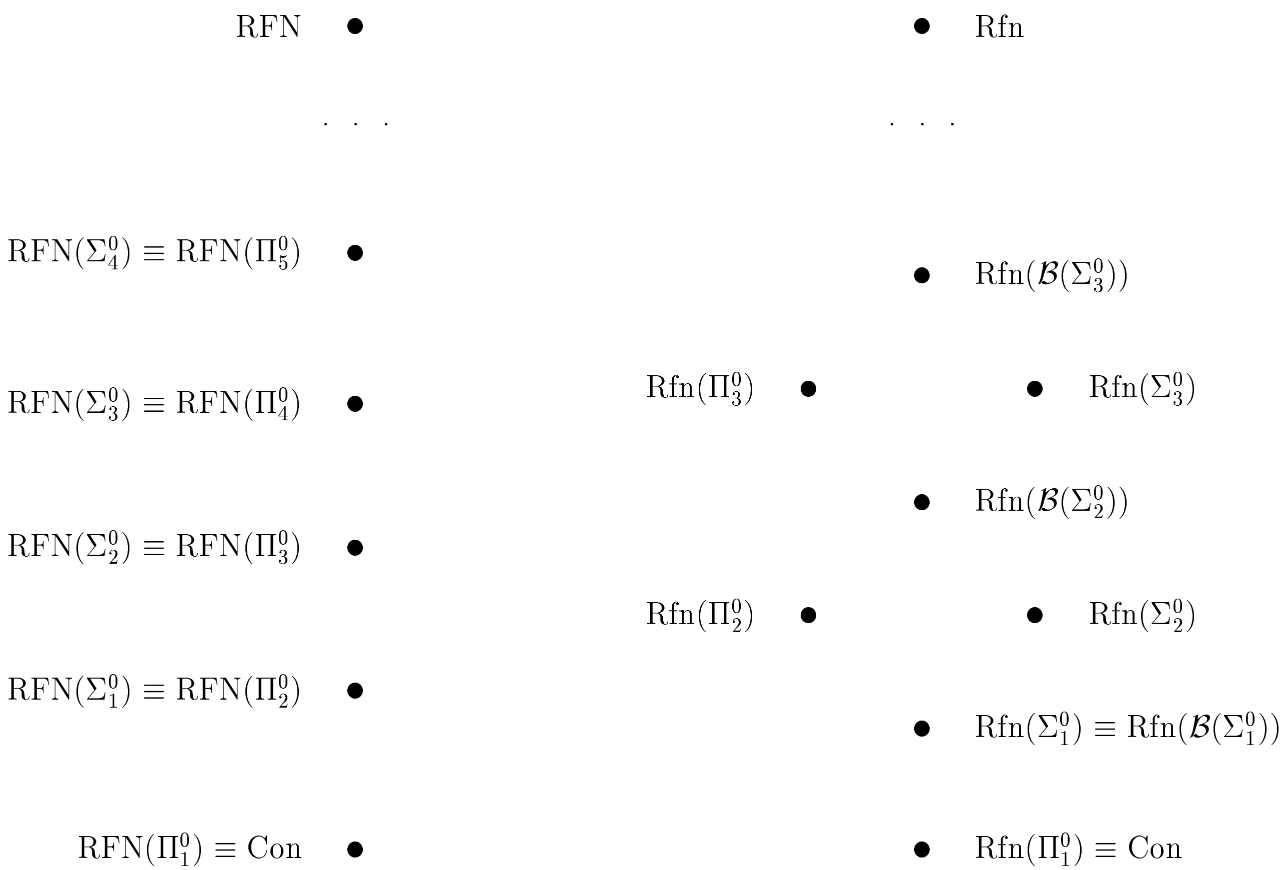

Figure 1: Hierarchies of partial reflection principles.

different manner. We mention them for the sake of comparison. Our results can be depicted by the diagram shown on Figure 1.

For the hierarchy of uniform reflection principles we have the following results.

1) $\mathrm{RFN}_T(\Sigma^0_n)$ is provably equivalent to $\mathrm{RFN}_T(\Pi^0_{n+1})$ over T. $\mathrm{RFN}_T(\Pi^0_1)$ is equivalent to Con_T, the consistency assertion for T (cf [8]). (Here and below we implicitly assume that $n \geq 1$.)

2) The schemata $\mathrm{RFN}_T(\Pi^0_n)$ and $\mathrm{RFN}_T(\Sigma^0_n)$ are finitely axiomatizable over T. In fact, the whole of $\mathrm{RFN}_T(\Pi^0_n)$ is equivalent to its particular instance

$$\forall x\, (\mathrm{Prov}_T(\ulcorner \mathrm{True}_{\Pi^0_n}(\dot{x})\urcorner) \rightarrow \mathrm{True}_{\Pi^0_n}(x)), \tag{1}$$

where $\mathrm{True}_{\Pi^0_n}(x)$ is the canonical truthdefinition for arithmetical Π^0_n formulas, i.e., a Π^0_n formula expressing the predicate "x is a Gödel number of a true Π^0_n sentence". Slightly abusing our notation, we shall also denote formula (1) by $\mathrm{RFN}_T(\Pi^0_n)$.

3) The theory $T + \mathrm{RFN}_T(\Pi^0_{n+1})$ proves the consistency of $T + \mathrm{RFN}_T(\Pi^0_n)$. Indeed, $\mathrm{RFN}_T(\Pi^0_n)$ is (equivalent to) a Π^0_n-sentence, hence $\mathrm{RFN}_T(\Pi^0_{n+1})$ proves

$$\mathrm{Prov}_T(\ulcorner \neg\mathrm{RFN}_T(\Pi^0_n)\urcorner) \rightarrow \neg\mathrm{RFN}_T(\Pi^0_n).$$

Trivially, $\mathrm{RFN}_T(\Pi^0_n)$ is contained in $\mathrm{RFN}_T(\Pi^0_{n+1})$, whence, by contraposition, $\mathrm{RFN}_T(\Pi^0_{n+1})$ implies $\neg\mathrm{Prov}_T(\ulcorner \neg\mathrm{RFN}_T(\Pi^0_n)\urcorner)$, that is, that $T + \mathrm{RFN}_T(\Pi^0_n)$ is consistent, q.e.d.

4) An immediate corollary of the above is the fact that the hierarchy of partial uniform reflection principles over T does not collapse (that is, all theories of the hierarchy are distinct) if and only if $T + \mathrm{RFN}_T$ is consistent. Moreover, if the hierarchy is proper, we have nonconservation of Π^0_1-sentences at each level of the hierarchy. (Clearly, soundness of T is sufficient for the theory $T + \mathrm{RFN}_T$ to be consistent. From the results of U.Schmerl (cf [7]) it follows that the consistency of $PA + \mathrm{RFN}_{PA}$ is equivalent to ϵ_1 times iterated consistency of PA, which provides a nice *necessary* and sufficient condition for the collapse of the hierarchy of partial uniform reflection principles for the case $T = PA$.)

For the hierarchy of partial local reflection principles the picture becomes quite different.

1) All the schemata $\mathrm{Rfn}_T(\Sigma^0_n)$ and $\mathrm{Rfn}_T(\Pi^0_n)$ are pairwise distinct, provided the theory $T + \mathrm{Rfn}_T$ is consistent. In particular, $\mathrm{Rfn}_T(\Sigma^0_n)$ is strictly contained in $\mathrm{Rfn}_T(\Pi^0_{n+1})$, and $\mathrm{Rfn}_T(\Sigma^0_n)$ and $\mathrm{Rfn}_T(\Pi^0_n)$ are incomparable for $n > 1$ (see Figure 1). (By a theorem of S.Goryachev (cf [4]), consistency of $T + \mathrm{Rfn}_T$ is equivalent to ω times iterated consistency of T. Observe that this ordinal bound is much smaller than ϵ_1, cf also [2].)

2) None of the schemata $\mathrm{Rfn}_T(\Sigma^0_n)$ and $\mathrm{Rfn}_T(\Pi^0_n)$ is finitely axiomatizable over T. (The only exception is the schema $\mathrm{Rfn}_T(\Pi^0_1)$, which is equivalent to consistency of T.) In fact, each of these theories is a reflexive extension of T (cf [1]).

3) Another corollary of Goryachev's theorem is the fact that all the restricted local reflection schemata, with the obvious exception of $\mathrm{Rfn}_T(\Pi^0_1)$, are mutually Π^0_1-conservative over each other, i.e., they prove the same Π^0_1 sentences. In particular, $\mathrm{Rfn}_T(\Pi^0_{n+1})$ does not prove the consistency of $T + \mathrm{Rfn}_T(\Pi^0_n)$.

Below we will show that an even stronger result holds: the full reflection schema Rfn_T is Π^0_n (resp., Σ^0_n) conservative over $T + \mathrm{Rfn}_T(\Pi^0_n)$ (resp., $T + \mathrm{Rfn}_T(\Sigma^0_n)$). This conservation result is optimal w.r.t. the arithmetical complexity in the sense that Σ^0_n (resp., Π^0_n) sentences are already not conserved.

Now we shall give the proofs of the facts mentioned in 1)–3) for local reflection principles. First of all, we show that restricted local reflection principles are related to each other in accordance with Figure 1. The relationships will follow from the next 2 lemmas. Everywhere below we assume T to be a primitively recursively axiomatized theory containing PRA and formulated in the language of PRA.

Lemma 1. *$\mathrm{Rfn}_T(\Sigma^0_1)$ is equivalent to $\mathrm{Rfn}_T(\mathcal{B}(\Sigma^0_1))$ over T.*

Proof. It is sufficient to show that $\mathrm{Rfn}_T(\Sigma^0_1)$ proves $\mathrm{Prov}_T(\ulcorner B \urcorner) \rightarrow B$ for any boolean combination of Σ^0_1-sentences B. Any such formula B can be equivalently written in the form

$$\bigwedge_{i=1}^{n} (\pi_i \vee \sigma_i),$$

for some sentences $\pi_i \in \Pi^0_1$ and $\sigma_i \in \Sigma^0_1$. Since the provability predicate $\mathrm{Prov}_T(\cdot)$ commutes with conjunction, it is sufficient to derive in $T + \mathrm{Rfn}_T(\Sigma^0_1)$ the formulas

$$\mathrm{Prov}_T(\ulcorner \pi_i \vee \sigma_i \urcorner) \rightarrow (\pi_i \vee \sigma_i),$$

for every i. We reason as follows:

(1)	$\mathrm{Prov}_T(\ulcorner \pi_i \vee \sigma_i \urcorner)$	(assumption)
(2)	$\neg \pi_i$	(assumption)
(3)	$\mathrm{Prov}_T(\ulcorner \neg \pi_i \urcorner)$	(from (2) by Σ^0_1-completeness)
(4)	$\mathrm{Prov}_T(\ulcorner \sigma_i \urcorner)$	(by (1),(3), and propositional logic inside $\mathrm{Prov}_T(\cdot)$)
(5)	σ_i	(from (4) by Σ^0_1-reflection)

Thus, we have shown that $\mathrm{Prov}_T(\ulcorner \pi_i \vee \sigma_i \urcorner)$ together with $\neg \pi_i$ implies σ_i, and this yields the required result by propositional logic, q.e.d.

Lemma 2. *For $n > 1$, neither of the schemata* $\mathrm{Rfn}_T(\Sigma^0_n)$ *and* $\mathrm{Rfn}_T(\Pi^0_n)$ *implies the other (over any theory T such that $T + \mathrm{Rfn}_T$ is consistent).*

Proof. Notice that, for $n > 1$, $\mathrm{Rfn}_T(\Sigma^0_n)$ is a primitive recursive set of Σ^0_n sentences consistent with T, and $\mathrm{Rfn}_T(\Pi^0_n)$ is a consistent p.r. set of Π^0_n sentences. Theorem 4 of Lindström [6] implies that if X is a p.r. set of Σ^0_n sentences and the theory $T + X$ is consistent, then there exists a *single* Σ^0_n sentence A such that $T + A$ is consistent and contains $T + X$. A similar result holds p.r. sets of Π^0_n-sentences.

Now let A be such a majorizing Σ^0_n sentence for $\mathrm{Rfn}_T(\Sigma^0_n)$. Clearly — and this was already noted by Kreisel and Lévy [5] — the formula A cannot consistently majorize $\mathrm{Rfn}_T(\Pi^0_n)$, for otherwise one would have

$$T + A \vdash \mathrm{Prov}_T(\ulcorner \neg A \urcorner) \to \neg A,$$

whence

$$T + A \vdash \neg \mathrm{Prov}_T(\ulcorner \neg A \urcorner),$$

and by Gödel's second incompleteness theorem $T + A$ would be inconsistent. It follows that $\mathrm{Rfn}_T(\Pi^0_n)$ is not contained in $\mathrm{Rfn}_T(\Sigma^0_n)$. The opposite noninclusion is proved symmetrically, q.e.d.

Corollary 1. *All inclusions shown on Figure 1 are strict.*

Proof. This is clear for the lowermost edge of the diagram, because $T + \mathrm{Rfn}_T(\Sigma^0_1)$ proves the consistency of $T + \mathrm{Con}_T$. Strictness of all other inclusions follows from the previous lemma, q.e.d.

A stronger form of the following lemma was proved in [1].

Lemma 3. *For $n \geq 1$, none of the schemata* $\mathrm{Rfn}_T(\Pi^0_{n+1})$ *and* $\mathrm{Rfn}_T(\Sigma^0_n)$ *is finitely axiomatizable over T (provided $T + \mathrm{Rfn}_T$ is consistent).*

Proof. By Goryachev's theorem each of these schemata is strong enough to prove the consistency of T together with an arbitrary finite number of instances of local reflection principle. In particular, $T + \mathrm{Rfn}_T(\Sigma^0_n)$ proves the consistency of each finitely axiomatized (over T) subtheory of itself. Gödel's second incompleteness theorem yields the result. Similar argument works for $\mathrm{Rfn}_T(\Pi^0_{n+1})$, q.e.d.

Theorem 1. *For $n > 1$,*

1. *The theories $T + \mathrm{Rfn}_T$ and $T + \mathrm{Rfn}_T(\Pi^0_n)$ prove the same Π^0_n-sentences.*
2. *The theories $T + \mathrm{Rfn}_T$ and $T + \mathrm{Rfn}_T(\Sigma^0_n)$ prove the same Σ^0_n-sentences.*
3. *$T + \mathrm{Rfn}_T$ and $T + \mathrm{Rfn}_T(\Sigma^0_1)$ prove the same $\mathcal{B}(\Sigma^0_1)$ sentences.*

Proof. Our proof borrows some ideas and results from provability logic. We refer the reader to the textbooks [3, 9] for an introduction into the subject. All the terminology left unexplained below can also be found there.

Let H_m denote the following propositional modal formula

$$\bigwedge_{i=1}^{m}(\Box p_i \to p_i).$$

Further, let p be a propositional variable distinct from all p_i's, and let the formulas P_k be defined inductively as follows:

$$P_0 := p; \qquad P_{k+1} := (P_k \wedge \Diamond P_k).$$

Everywhere below GL denotes the basic provability logic of Gödel and Löb (C.Smoryński calls it PrL).

Lemma 4. *For all m, GL proves*

$$\Diamond P_m \to \Diamond(p \wedge H_m). \tag{2}$$

Proof. First of all, an easy induction on m shows that

$$GL \vdash P_m \to p, \tag{3}$$

for all m. To demonstrate (2) we use a Kripke model argument. Consider an arbitrary finite irreflexive treelike Kripke model for GL, and let x be a node of this model where the formula $\Diamond P_m$ is forced. By the construction of P_m, there is an increasing chain of $m+1$ nodes above x

$$x \prec x_m \prec x_{m-1} \prec \ldots \prec x_0$$

such that, for all i, $x_i \Vdash P_i$. From (3) we conclude that $x_i \Vdash p$ for all i.

Now we look at the formula H_m and notice that every conjunct $\Box p_i \to p_i$ can be false at no more than 1 node of the chain $x_m \prec x_{m-1} \prec \ldots \prec x_0$. By Pigeon-hole Principle there exists a node among the x_i's where H_m is true. It follows that $\Diamond(p \wedge H_m)$ must be true at x, q.e.d.

Denote $Q_m := \neg P_m$.

Lemma 5. *For all m, GL proves*

$$\Box(H_{m+1} \to \neg p) \to \Box(\bigwedge_{i=0}^{m}(\Box Q_i \to Q_i) \to \neg p). \tag{4}$$

Proof. Clearly, within GL, the formula $\bigwedge_{i=0}^{m}(\Box Q_i \to Q_i)$ implies $\bigwedge_{i=0}^{m}(P_i \to \Diamond P_i)$. On the other hand, by induction on m it is easy to show that $p \wedge \bigwedge_{i=0}^{m}(P_i \to \Diamond P_i)$ implies P_{m+1}, so we have:

$$GL \vdash \neg P_{m+1} \to (\bigwedge_{i=0}^{m}(\Box Q_i \to Q_i) \to \neg p). \qquad (5)$$

Now we reason inside GL as follows. From $\Box(H_{m+1} \to \neg p)$ infer $\Box\neg(p \wedge H_{m+1})$ and $\Box\neg P_{m+1}$ (by (2)). By (5) we obtain $\Box(\bigwedge_{i=0}^{m}(\Box Q_i \to Q_i) \to \neg p)$, q.e.d.

Proof of Theorem 1. According to the standard arithmetical interpretation of provability logic the modal formula (4) should be read as follows: if an arithmetical formula $\neg p^*$ is provable in T from arbitrary $m+1$ instances of local reflection schema for T, then it can already be proved from $m+1$ instances of local reflection for specific arithmetic formulas Q_i^*. The formulas Q_i^* are inductively defined as follows (see the dual definition of P_i):

$$Q_0^* := \neg p^*; \qquad Q_{i+1}^* := (Q_i^* \vee \mathrm{Prov}_T(\ulcorner Q_i^* \urcorner)).$$

Observe that, if $\neg p^* \in \Sigma_n^0$, then, for all i, $Q_i^* \in \Sigma_n^0$. So, Σ_n^0 consequences of Rfn_T are provable in $\mathrm{Rfn}_T(\Sigma_n^0)$. Similarly, for $n > 1$, Π_n^0 consequences of Rfn_T are provable in $\mathrm{Rfn}_T(\Pi_n^0)$, and $\mathcal{B}(\Sigma_1^0)$ consequences of Rfn_T are provable in $\mathrm{Rfn}_T(\mathcal{B}(\Sigma_1^0))$, which is equivalent to $\mathrm{Rfn}_T(\Sigma_1^0)$ by Lemma 1, q.e.d.

The following lemma shows that conservation results of Theorem 1 are optimal w.r.t. the arithmetical complexity.

Lemma 6. *For $n > 1$, there is a Σ_n^0 (resp., Π_n^0) sentence provable in $T + \mathrm{Rfn}_T$ but not in $T + \mathrm{Rfn}_T(\Pi_n^0)$ (resp., $T + \mathrm{Rfn}_T(\Sigma_n^0)$), provided $T + \mathrm{Rfn}_T$ is consistent.*

Proof: Follows from Lemma 2.

Acknowledgement. This paper was inspired by a discussion held at the colloquium "Mathematische Logik" at Mathematisches Forschungsinstitut Oberwolfach, April 2–8, 1995.

References

[1] L.D. Beklemishev. Bimodal logics for extensions of arithmetical theories. Prepublication Series LCS 7, Department of Math. Logic, Steklov Math. Institute, Moscow, November 1992. To appear in JSL.

[2] L.D. Beklemishev. Iterated local reflection vs. iterated consistency. Logic Group Preprint Series 118, University of Utrecht, 1994. To appear in *Annals of Pure and Applied Logic*.

[3] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

[4] S. Goryachev. On interpretability of some extensions of arithmetic. *Mat. Zametki*, 40:561–572, 1986. In Russian.

[5] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:97–142, 1968.

[6] P. Lindström. On partially conservative sentences and interpretability. *Proceedings of the AMS*, 91(3):436–443, 1984.

[7] U.R. Schmerl. A fine structure generated by reflection formulas over Primitive Recursive Arithmetic. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium'78*, pages 335–350. North Holland, Amsterdam, 1979.

[8] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North Holland, Amsterdam, 1977.

[9] C. Smoryński. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin-Heidelberg-New York, 1985.