ON REFLECTION PRINCIPLES

By

Akito Tsuboi

Introduction.

In this paper, we shall consider various forms of reflection principles for 1-st order theories containing arithmetic. If a 1-st order theory $T$ contains arithmetic, we can express various notions concerning $T$ within $T$ itself, by using the coding method developed by K. Gödel. K. Gödel assigned each formula $\phi$ in the language of $T$ a number $^{\gamma}\phi$ (the Gödel number of $\phi$), but our method is slightly different.

We assume that variables, individual constants, relation symbols and function symbols are numbers, and logical symbols ($*_{X}, *_{V}, \neg_{*}, *, \forall_{*}, \exists_{*}$) are operations on numbers. Under these assumptions, a formula $\alpha(x)$ in the language of $S$ numerates $A$ in $S$ if, for any $n \in \omega$,

$$n \in A \text{ iff } S \text{ proves } \alpha(n),$$

where $\bar{n}$ denotes the $n$-th numeral, i.e., the term of $S$ which expresses the number $n$. In this case, we call this $\alpha$ a numeration of $A$ in $S$. If $\alpha$ numerates $A$ in $S$ and $\neg_{A}$ numerates $\omega\setminus A$ in $S$, we say $\alpha$ binumerates $A$ in $S$, and $\alpha$ is called a binumeration of $A$ in $S$.

Let $A=\{n_{1}, \cdots, n_{m}\}$ be a subset of $\omega$. Then $[A]$ denotes the formula $x=\bar{n}_{1} \cdots x=\bar{n}_{m}$. Clearly $[A]$ binumerates $A$ in any theory $S$ which contains arithmetic.

If a binumeration $\tau$ of a theory $T$ in a theory $S$ is given, we can construct a provability formula $Pr_{\tau}(x)$ whose intuitive meaning is that a formula $x$ is provable in $T$. The reader should note that this $Pr_{\tau}$ cannot be uniquely determined by $T$, but is determined by $\tau$. (The explicit definition of $Pr_{\tau}$ can be found in p. 59 of [1].)

Using this $Pr_{\tau}$, we define the $\tau$-reflection principle $Rf_{\tau}(\tau)$ and the $\tau$-reflection principle $Rf_{\tau_{A}}(\tau)$ based on $A$:

Received October 9, 1981. Revised March 26, 1982.
\[ Rfn(\tau) = \{ Pr_\tau(\sigma) \mid \sigma \in \text{Sent}_T \}, \]
\[ Rfn_A(\tau) = \{ Pr_\tau(\sigma) \mid \sigma \in A \cap \text{Sent}_T \}, \]
where \( \text{Sent}_T \) is the set of all sentences in the language of \( T \).

A formula \( \phi \) is said to be a \( \Sigma_n \)-formula \( (\Pi_n \text{-formula}) \) if \( \phi \) has the form \( Q_1x_1 \cdots Q_nx_n \psi(x_1, \ldots, x_n) \) for some quantifier bounded formula \( \phi \), where \( Q_1 = \exists(\forall) \) and the quantifiers alternate in type. The set of all \( \Sigma_n \)-formulas \( (\Pi_n \text{-formulas}) \) is denoted by \( \Sigma_n \) \( (\Pi_n) \).

Let \( T \) be a 1-st order theory containing arithmetic. We say \( T \) is \( n \)-consistent if the following two conditions are not simultaneously satisfied for any \( \Pi_n \)-formula \( \phi \):

i) \( T \) proves \( \exists x \phi(x) \),

ii) \( T \) proves \( \neg \phi(\bar{m}) \) for all \( m \in \omega \).

If \( T \) is \( n \)-consistent for all \( n \in \omega \), we say \( T \) is \( \omega \)-consistent. If \( T \) proves \( \text{Con}_{T_\emptyset} \) \( (= \neg Pr_{T_\emptyset}(\overline{\emptyset}) \) for all finite subtheories \( T_\emptyset \) of \( T \), \( T \) is said to he reflexive. If each extension \( T^* \) of \( T \) with the same language is reflexive, we say \( T \) is essentially reflexive. We next define a more complicated notion \( A \)-reflexiveness. Let \( A \) be a set of sentences. We say \( T \) is \( A \)-reflexive, if there exist a truth definition \( Tr_A(x) \) for \( A \) in \( T \) and a numeration \( \alpha(x) \) of \( A \) in \( T \) for which \( T \) proves \( \forall x(\alpha(x) \land \text{Sent}(x) \land Pr_{T_\emptyset}(x) \rightarrow Tr_A(x)) \) for all finite subtheories \( T_\emptyset \) of \( T \), where \( \text{Sent}(x) \) is a formula which expresses that \( x \) is a sentence. (See Definition 1.2 and 1.3 for reference.)

For three sets \( A, B \) and \( C \) of sentences, we put:

\[ A \subseteq_B C \text{ iff each sentence in } A \text{ is provable in } B \cup C, \]
\[ A =_B C \text{ iff } A \subseteq_B C \text{ and } C \subseteq_B A, \]
\[ A \equiv_B C \text{ iff } A \subseteq_B C \text{ and } A \neq_B C. \]

In case \( B \) is the empty set, we usually omit \( B \) in the above definitions. In what follows, we say \( S \) is a subtheory of \( T \) if \( S \subseteq T \) holds in this sense. If \( T \) is a theory and \( A \) is a set of sentences in the language of \( T \), then we put \( T - A = \{ \phi \mid \phi \text{ is equivalent to some } \phi \in A \text{ in } T \}. \)

It is now possible to state the main theorems of this paper.

**Theorem 1.** Suppose that \( A \) is a set of sentences. If \( T \) is an \( A \)-reflexive theory with a binumeration \( \tau \) of \( T \) in \( T \), then we can effectively construct a binumeration \( \tau' \) of \( T \) in \( T \) for which \( T \) proves each member of \( Rfn_A(\tau') \).

**Theorem 2.** Suppose that \( T \) is a recursively enumerable theory (r.e. theory)
and $S$ is a subtheory of $T$. If $\tau$ binumerates $T$ in $S$, then $Rfn(\tau)\setminus Rfn_{T-\omega}(\tau) = \tau Rfn(\tau)$.

**Theorem 3.** Suppose that $T$ is an $\omega$-consistent and essentially reflexive theory and $S$ is a subtheory of $T$. If $\tau$ binumerates $T$ in $S$, we can effectively construct binumerations $\tau_1$ and $\tau_2$ of $T$ in $S$ for which $Rfn(\tau_1) = \tau Rfn(\tau_2)$.

Theorem 1, which appears in § 2, is closely related to Theorem 5.9 of [1]. Theorem 2 shows that the strength of $Rfn(\tau)$ does not change even if the lower part of it is taken away from it. Theorem 2 also appears in § 2. Theorem 3, which appears in § 3, is an analogy of Theorem 7.4 and 7.5 of [1] and shows that the choice of numerations must be done very carefully.

The reader who is accustomed to the coding method can skip § 1 and may refer to it as occasion demands.

The author wishes to express his heartfelt thanks to Prof. N. Motohashi for a number of helpful suggestions.

§ 1. Preliminaries.

Notations, definitions and conventions in this paper largely correspond with those of [1]. Especially, we assume that a formula $\phi$ itself is a number and we do not use a notation $\tau^\phi$ (the Gödel number of $\phi$).

For simplicity, we say $T$ is a theory when $T$ is a consistent 1-st order theory containing PA (Peano arithmetic). We use $T$, $S$ and $T_i$ ($i=1, 2, \cdots$) as syntactic variables ranging over theories, and usually assume that $S$ is a subtheory of $T$.

It will be convenient to assume that for every theory $T$, $L_T$ (the language of $T$) has all the symbols for p.r. functions (primitive recursive functions) and $T$ contains all the defining axioms for p.r. functions. The function symbol associated with a p.r. function $f$ is denoted by $\overline{f}$ or $f$. But symbols which are used very often will be used without dots. For example, we write $\nu_r x$ and $\text{num}_x$ (the $x$-th variable and the $x$-th numeral, respectively) instead of writing $\nu_r x$ and $\text{num}_x$. $Tr(x)$, $Fm(x)$, $\text{Sent}(x)$, $Prf_r(x, y)$, $Prf_r(x)$ and $\text{Con}_r$ are formulas whose intuitive meanings are "$x$ is a term", "$x$ is a formula", "$x$ is a sentence", "$y$ is a proof of $x$ in a theory with a numeration $\tau$" and "a theory with a numeration $\tau$ is consistent", respectively.

**Convention.** Let $\alpha$ be a formula with a free variable $x$ and $A$ be a set of formulas. We say $\alpha$ numerates $A$ in a theory $T$, if $\alpha$ numerates $A$ in $T$ in the
usual sense and $T$ proves $\forall x(\alpha(x) \rightarrow Fm(x))$. (See Introduction for reference.)

The above convention is trivial: If $\alpha$ numerates $A$ in $T$ in the usual sense and $T$ does not prove $\forall x(\alpha(x) \rightarrow Fm(x))$, then we can define $\alpha'$ as $\alpha(x).Fm(x)$, and $\alpha'$ will numerate $A$ in $T$ in the above sense.

In the following, we state some definitions which do not appear in [1].

1.1. Definition. Let $T$ be a theory and $S$ a subtheory of $T$. Then we put:

$$\text{Bin}(T, S) = \{ \tau \mid \tau \text{ binumerates } T \text{ in } S \}.$$

1.2. Definition. Let $T$ be a theory and $A$ a set of formulas in the language of $T$. We say a formula $Tr_A(x)$ is a truth definition for $A$ in $T$, if the following is satisfied:

$$T \vdash T_{Tr_A}(\phi(x_1, \ldots, x_n)) \leftrightarrow \phi(x_1, \ldots, x_n) \text{ for all } \phi(x_1, \ldots, x_n) \in A,$$

where $\bar{\phi}(x_1, \ldots, x_n) = \text{sub}(\phi; x_1, \ldots, \bar{x}_n/nm_{x_1}, \ldots, nm_{x_n})$, i.e., the sentence obtained from $\phi$ by substituting $nm_{x_1}, \ldots, nm_{x_n}$ for its free variables $x_1, \ldots, x_n$.

1.3. Definition. Let $T$ be a theory and $A$ a set of sentences in the language of $T$. Then we say

i) $T$ is reflexive if $T \vdash \text{Con}_{T_{\phi}}$ for every finite subtheory $T_{\phi}$ of $T$,

ii) $T$ is essentially reflexive if every extension $T^*$ of $T$ in the same language as $T$ is reflexive,

iii) $T$ is $A$-reflexive if there exist a numeration $\alpha$ of $A$ in $T$ and a truth definition $T_{Tr_A}$ for $A$ in $T$ for which

$$T \vdash \forall x(\alpha(x) \land \text{Sent}(x) \land Pr_{T_{\phi}}(x) \rightarrow T_{Tr_A}(x))$$

holds for every finite subtheory $T_{\phi}$ of $T$.

1.4. Corollary. The following i), ii) and iii) are equivalent:

i) $T$ is essentially reflexive,

ii) $T \vdash Pr_{T_{\phi}}(\phi) \rightarrow \phi$ for every $\tau \in \text{Bin}(T, T)$, $\phi \in \text{Sent}_{T}$, $n \in \omega$, (where, of course, $\tau \downarrow \bar{n}$ is an abbreviation for $\tau(x), x \leq \bar{n}$.)

iii) $T$ is $\{\phi\}$-reflexive for every $\phi \in \text{Sent}_{T}$.

Since there is no truth definition for all sentences, a reflection principle cannot be formulated in a single sentence. Although there are many versions of a reflection principle, we restrict our attention to the following two types.
1.5. **Definition.** Let $T$ and $S$ be theories with $S \subseteq T$ and let $\tau$ be a formula which binumerates $T$ in $S$. Then:

i) Local Reflection Principle;

\[ Rfn(\tau) = \{ \Pr(\phi) \rightarrow \phi | \phi \in \text{Sent}_T \}, \]

\[ Rfn_A(\tau) = \{ \Pr(\phi) \rightarrow \phi | \phi \in A \setminus \text{Sent}_T \}, \]

ii) Uniform Reflection Principle;

\[ RFN(\tau) = \{ \forall x \in \Sigma_n \cup \Pi_n \forall y (\Pr(x^*) \rightarrow Trn(x^*)) | n \in \omega \}, \]

where $Trn$ is the standard truth definition for $\Sigma_n \cup \Pi_n$ and $x^*$ denotes the sentence obtained from $x$ by substituting $nm_{(y), nm_{(y)}}, \ldots$ for its free variables.

Does $T$ remain consistent when a reflection principle is added to it? The following theorem gives us a partial solution.

1.6. **Theorem.** (Refinement of Theorems 20 and 24 of [4]) Let $T$, $S$ and $\tau$ be as above. Then:

i) If $\tau \in \Sigma_1$ and $T$ is 1-consistent, then $T \cup Rfn(\tau)$ is 1-consistent,

ii) If $\tau \in \Sigma_n$ and $T$ is $n$-consistent, then $T \cup Rfn(\tau)$ is 2-consistent. ($n = 2, 3, \ldots$),

iii) If $T$ is $\omega$-consistent, then $T \cup RFN(\tau)$ is 2-consistent.

**Remark.** T. Miyatake showed that if $\tau \in \Sigma_1$, then the converse of iii) also holds. If $\tau \in \Sigma_1$, the 2-consistency is not enough, but the weak converse of iii) holds, and it can be stated as follows: if $\tau \in \Sigma_n$ and $T \cup RFN(\tau)$ is $n+1$-consistent, then $T$ is $\omega$-consistent. It is not hard to give an example of $T$ for which $T \cup Rfn(\tau)$ is inconsistent. The reader may refer to [4] for this purpose.

§ 2. **Hierarchy Considerations.**

By Gödel's Second Incompleteness Theorem, if $\tau \in Bin(T, T)$ is a $\Sigma_1$-formula, $Con_\tau$ cannot be proved in $T$. S. Feferman, however, in [1] showed that in case $T$ is reflexive, we can choose $\tau \in Bin(T, T)$ for which $PA$ proves $Con_\tau$. Since $Con_\tau$ and $Rfn_{\Pi_1}(\tau)$ are equivalent over $T$, we can also prove all elements of $Rfn_{\Pi_1}(\tau)$ in $T$ for the above $\tau$. The following theorem is a generalization of this fact.

2.1. **Theorem.** Suppose that $A$ is a set of sentences. If $T$ is an $A$-reflexive theory with a binumeration $\tau \in Bin(T, T)$, then we can effectively construct from $\tau$ a binumeration $\tau' \in Bin(T, T)$ for which $T$ prove each element of $Rfn_A(\tau')$. 

Proof. Since $T$ is $A$-reflexive, we can choose a numeration $\alpha(x)$ of $A$ in $T$ and a truth definition $Tr_A(x)$ for $A$ in $T$ such that

$$T \vdash \forall x (\alpha(x) \land Sent(x) \land Pr_{\alpha}(x) \rightarrow Tr_A(x))$$

for all $n \in \omega$.

Set $\beta(x) = \forall y (\alpha(y) \land Sent(y) \land Pr_{\alpha}(y) \rightarrow Tr_A(y))$, $\tau'(x) = \tau(x)$, $\forall y \leq \beta(y)$. We prove that this $\tau'$ has the desired properties. Since $\tau' \in Bin(T, T)$ is easily obtained from the assumptions, we have only to show that $T$ proves each element of $Rfn_n(\tau')$. First note that $T \vdash Pr_{\tau}(x) \rightarrow \exists y Pr_{\tau + y}(x)$, then

$$T \vdash \neg \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau}(x) \rightarrow Tr_A(x)) \rightarrow \exists y (\neg \beta(y)).$$

Now, by the assumption, $T \vdash \beta(0)$, hence

$$T \vdash \exists y (\neg \beta(y)) \rightarrow \exists y (\neg \beta(y') \land \forall z \leq \beta(y'))$$

$$\rightarrow \exists y (\beta(y') \land \exists x (\tau(x) \land \forall z \leq y \rightarrow \tau'(x)))$$

$$\rightarrow \exists y (\forall z (\alpha(z) \land Sent(z) \land Pr_{\tau'}(z) \rightarrow Tr_A(z)).$$

Thus we have

$$T \vdash \neg \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau}(x) \rightarrow Tr_A(x))$$

$$\rightarrow \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau'}(x) \rightarrow Tr_A(x)).$$  \hspace{1cm} (1)

On the other hand, by the definition of $\tau'$,

$$T \vdash \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau}(x) \rightarrow Tr_A(x))$$

$$\rightarrow \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau'}(x) \rightarrow Tr_A(x)).$$  \hspace{1cm} (2)

Combining (1) and (2), we have

$$T \vdash \forall x (\alpha(x) \land Sent(x) \land Pr_{\tau}(x) \rightarrow Tr_A(x)).$$

So

$$T \vdash Pr_{\tau}(\bar{\phi}) \rightarrow \phi \quad \text{for all } \phi \in A \setminus Sent_T,$$

as desired. $\square$

2.2. Corollary. Suppose that $T$ is an r.e. theory with the same language as PA. Then there is a theory $T^*$ with $T^* = T$, and for each $n \in \omega$, there is a $\tau_n \in Bin(T^*, T^*)$ such that $T^*$ prove each element of $Rfn_{\Sigma_n \cup \Pi_n}(\tau_n)$.

Proof. By Theorem 4.13 of [1], there is a theory $T^*$ with $T^* = T$, and there is a $\tau \in Bin(T^*, T^*)$. So it is sufficient to prove that

$$T \vdash \forall x \in \Sigma_n \cup \Pi_n \forall y (Pr_{\tau + n}(x) \rightarrow Tr_n(x))$$

for all $m, n \in \omega$.

By formalizing a proof of the soundness of a 1-st order logic, we have
Let \( m, n \in \omega \) be given. If \( \tau \vdash \bar{m} \) is equivalent to \( \bigvee_{i \leq j} x_i \), then, for sufficiently large \( j \geq n \),

\[
T \vdash \forall x \in \Sigma_n \cup \Pi_n \forall y(Pr_{\tau \vdash \bar{m}}(\bigwedge_i x_i \rightarrow y) \rightarrow Tr_j(\bigwedge_i x_i \rightarrow y)),
\]

\[
T \vdash \forall x \in \Sigma_n \cup \Pi_n \forall y(Pr_{\tau \vdash \bar{m}}(x^*) \rightarrow (\bigwedge_i x_i \rightarrow Tr_j(x^*))),
\]

\[
T \vdash \forall x \in \Sigma_n \cup \Pi_n \forall y(Pr_{\tau \vdash \bar{m}}(x^*) \rightarrow Tr_j(x^*)).
\]

Since \( Tr_n \) and \( Tr_j \) are standard ones, we have

\[
T \vdash \forall x \in \Sigma_n \cup \Pi_n \forall y(Tr_j(x^*) \rightarrow Tr_n(x^*))
\]

which completes our proof. \( \square \)

If \( T = PA \), we don't have to choose \( T^* \) as in the above corollary. So the following holds:

2.3. **Corollary.** For each \( n \in \omega \), there is a \( \pi_n \in Bin(PA, PA) \) for which \( PA \) proves each element of \( Rfn_{\Sigma_n \cup \Pi_n}^{\pi_n} \).

Now, we take another side view of the lower part of a reflection principle w.r. to the formula hierarchy.

2.4. **Theorem.** Suppose that \( T \) is an r.e. theory and \( S \) is a subtheory of \( T \). Then, for each \( \tau \in Bin(T, S) \) and \( n \in \omega \),

\[
Rfn(\tau) \setminus Rfn_{\tau \vdash \neg (\Sigma_n \cup \Pi_n)}(\tau) = T Rfn(\tau).
\]

To prove Theorem 2.4 we need some lemmas.

2.5. **Lemma (Kent).** If \( T^* \) is a consistent extension of an r.e. theory \( T \), obtained by the addition of axioms in \( \Sigma_n \cup \Pi_n \), in which each sentence of \( \Sigma_n \cup \Pi_n \) is decidable, then \( T^* \) is incomplete.

**Proof.** See Theorem 3 of [3].

2.6. **Lemma.** Suppose that \( T \) is an r.e. theory and \( \phi_0, \phi_1 \in \text{Sent}_T \). If \( \phi_0 \) and \( \phi_1 \) satisfy

\[
T \vdash \phi_0 \rightarrow \phi_1 \quad \& \quad T \vdash \phi_1 \rightarrow \phi_0,
\]

then, for each \( n \in \omega \), there is a \( \chi_n \in \text{Sent}_T \) for which

\[
T \vdash \phi_0 \rightarrow \phi_n \quad \& \quad T \vdash \chi_n \rightarrow \phi_1 \quad \& \quad \chi_n \in T \setminus (\Sigma_n \cup \Pi_n).
\]
Proof. By way of a contradiction, suppose that for an arbitrary \( \phi \in \text{Sent}_T \),
\[
T \vdash \neg \phi \rightarrow \phi \quad \& \quad T \vdash \phi \rightarrow \phi_1 \quad \text{implies} \quad \phi \in T - (\Sigma_n \cup \Pi_n) .
\] (1)
For each \( \phi \in \text{Sent}_T \), \( (\phi \lor \phi) \land \phi_1 \) satisfies the left side of (1). Thus, for any \( \phi \), there is a \( \sigma \in \Sigma_n \cup \Pi_n \) such that \( T \) proves \( (\phi \lor \phi) \land \phi_1 \rightarrow \sigma \), i.e.,
\[
T \cup \{ \neg \phi_0, \ \phi_1 \} \vdash \phi \rightarrow \sigma .
\] (2)
Adding \( \Sigma_n \cup \Pi_n \)-sentences to consistent \( T \cup \{ \neg \phi_0, \ \phi_1 \} \), we can construct a consistent theory \( T^* \) which is complete for \( \Sigma_n \cup \Pi_n \)-sentences. But (2) holds, therefore \( T^* \) must be complete. This contradicts the assertion of Lemma 2.5. \( \square \)

Now we can prove Theorem 2.4.

Proof of Theorem 2.4. It is sufficient to show that
\[
T \cup \text{Rfn}(\tau) \backslash \{ \text{Rfn}_T - \text{Sent}_{\Sigma_n \cup \Pi_n} \}(\tau) \vdash \text{Pr}_T (\phi) \rightarrow \phi ,
\]
for \( \phi \in \Sigma_n \cup \Pi_n \) such that \( T \vdash \phi \). Fix such a sentence \( \phi \in \Sigma_n \cup \Pi_n \). By Lemma 2.6, there is a \( \phi_1 \) for which
\[
T \vdash \phi \rightarrow \phi_1 \quad \& \quad \phi_1 \in T - (\Sigma_n \cup \Pi_n) . \quad (3)
\]
This \( \phi_1 \) is unprovable in \( T \). Hence, using Lemma 2.6 again, we can find \( \phi_2 \) for which
\[
T \vdash \phi \rightarrow \phi_2 \quad \& \quad T \vdash \phi_2 \rightarrow \phi_1 \rightarrow \phi \quad \& \quad \phi_2 \in T - (\Sigma_n \cup \Pi_n) . \quad (4)
\]
Combining (3) and (4) yields
\[
\phi_1, \ \phi_2 \in T - (\Sigma_n \cup \Pi_n) \quad \& \quad T \vdash \phi \rightarrow \phi_1 \land \phi_2 . \quad \ldots \quad (5)
\]
Therefore, we have
\[
(\text{Pr}_T (\phi_1) \rightarrow \phi_1) , \quad (\text{Pr}_T (\phi_2) \rightarrow \phi_2) \in \text{Rfn}(\tau) \backslash \{ \text{Rfn}_T - \text{Sent}_{\Sigma_n \cup \Pi_n} \}(\tau)
\]
\[
\& \quad T \vdash (\text{Pr}_T (\phi_1) \rightarrow \phi_1) \land (\text{Pr}_T (\phi_2) \rightarrow \phi_2) \rightarrow (\text{Pr}_T (\phi) \rightarrow \phi) ,
\]
as desired. \( \square \)

The following is an easy consequence of the above theorem, and we can safely leave its proof to the reader.

2.7. Corollary. Suppose that \( T \) is an r.e. theory and \( S \) is a subtheory of \( T \). Suppose that \( \tau \) is a binumeration of \( T \) in \( S \). Then, for each finite subset \( A \) of \( \text{Rfn}(\tau) \),
\[
\text{Rfn}(\tau) \backslash (T - A) = \tau \text{Rfn}(\tau) .
\]
§ 3. The Ordering of The \( \tau \)-Reflection Principles.

So far, we have investigated the behavior of the lower parts of the \( \tau \)-reflection principles (w.r. the formula hierarchy). In this section, we compare the strength of the whole \( Rfn(\tau) \)'s as new axioms of \( T \), and show that for a fixed \( \omega \)-consistent and essentially reflexive theory \( T \), there are no maximal elements and no minimal elements in \( \{ Rfn(\tau) | \tau \in Bin(T, S) \} \) w.r. to \( \sqsubseteq_T \).

3.1. Theorem. Suppose that \( T \) is an essentially reflexive theory. Then for each \( \tau \in Bin(T, S) \), we can effectively construct a \( \tau' \in Bin(T, S) \) for which

\[
Rfn(\tau') \sqsubseteq_T Rfn(\tau).
\]

Proof. A simple diagonal argument shows that there is a \(( Pr_T(\bar{\sigma}) \to \sigma) \in Rfn(\tau) \) which is not provable in \( T \). Using this \( Pr_T(\bar{\sigma}) \to \sigma \), we define \( \alpha(x) \) by

\[
\alpha(x) = \tau(x) \forall x = \neg(Pr_T(\bar{\sigma}) \to \sigma).
\]

For each formula \( \gamma(x) \), define \( f_T(m) \) by

\[
f_T(m) = \bigwedge_{\phi \in m} (Pr_T(\bar{\phi}) \to \phi).
\]

Using these, we define a diagonal sentence \( \phi_n \) such that

\[
PA \vdash \phi_n \leftrightarrow \forall x(Pr_\alpha(\bar{\phi}_n, x) \to \neg f_{T, x}(n)).
\]

\( \phi_n \) can be constructed effectively from \( n \) (in fact primitive recursively from \( n \)). Hence, there is a corresponding p.r. function symbol \( \bar{\phi} \) such that \( PA \vdash \bar{\phi}_n = \bar{\phi}(\bar{n}) \).

Now define \( \tau'(x) \) by

\[
\tau'(x) = \tau(x) \forall y, z \leq x \neg Pr_\alpha(\bar{\phi}(y), z).
\]

First we prove that

\[
T \vdash \neg(Pr_T(\bar{\sigma}) \to \sigma) \to \phi_n \quad \text{for all} \quad n \in \omega.
\] (1)

Assume that \( T \vdash \neg(Pr_T(\bar{\sigma}) \to \sigma) \to \phi_n \), then \( T \vdash Pr_\alpha(\bar{\phi}_n, \bar{m}) \) for some \( m \in \omega \). Hence, using the definition of \( \phi_n \), we have \( T \vdash \neg(Pr_T(\bar{\sigma}) \to \sigma) \to f_{T, m}(n) \). On the other hand, by the essential reflexiveness of \( T \), \( T \vdash f_{T, m}(n) \). So \( T \vdash Pr_T(\bar{\sigma}) \to \sigma \). But this is a contradiction, which leads us to conclude (1). Next we prove that

\[
T \vdash \neg \phi_n \to f_{T, m}(n) \quad \text{for all} \quad n \in \omega.
\]

Since we can assume \( T \vdash \forall y(\phi(y) > y) \), \( T \vdash \forall y \forall x \forall z \geq x \neg Pr_\alpha(\bar{\phi}(y), z) \). So we have

\[
T \vdash \forall x(\tau'(x) \to \tau(x), \forall y \forall z \leq x \neg Pr_\alpha(\bar{\phi}_n, z)) \quad \text{for all} \quad n \in \omega.
\]

Now, using the definition of \( \phi_n \), we have
Theorem 3.1 asserts that \( \{Rfn(\tau') | \tau \in Bin(T, S) \} \) has no minimal elements w.r. to \( \equiv_T \), if \( T \) is essentially reflexive. Maximal elements also do not exist, if \( T \) is \( \omega \)-consistent.

3.2. Theorem. Suppose that \( T \) is an \( \omega \)-consistent theory. Then, for each \( \tau \in Bin(T, S) \), we can effectively construct a \( \tau' \in Bin(T, S) \) for which

\[
Rfn(\tau') \equiv_T Rfn(\tau').
\]

Proof. Set \( T' = T \cup Rfn(\tau) \). Then Theorem 1.6 guarantees the consistency of \( T' \). Let \( \beta'(x) \) be a \( \Sigma_0 \)-formula which binumerates \( Rfn(\tau) \) in \( PA \). Using this \( \beta' \) define \( \beta(x) \) by

\[
\beta(x) = \tau(x) \lor \beta'(x).
\]

Clearly, \( \beta \) is a binumeration of \( T' \) in \( S \). By Gödel's theorem,

\[
T' \vdash \nu_{\beta},
\]

where \( \nu_{\beta} \) is a fixed point of \( \neg Prf_{\beta}(x) \). Next define \( \tau'(x) \) by

\[
\tau'(x) = \tau(x) \lor Fm(x) \land \exists y < x Prf_{\beta}(\overline{\nu_{\beta}}, y).
\]

Then \( \tau' \) is a binumeration of \( T \) in \( S \). Since

\[
PA \vdash \neg \nu_{\beta} \rightarrow \exists x Prf_{\beta}(\overline{\nu_{\beta}}, x)
\]

\[
\rightarrow \exists y \exists x \neg (\neg (\nu_{\beta} = x) Prf_{\beta}(\overline{\nu_{\beta}}, x))
\]

\[
\rightarrow \exists y \tau'(\neg (\nu_{\beta} = x))
\]

\[
\rightarrow \neg Con_{\tau'},
\]

we have \( PA \vdash Con_{\tau'}, \neg \nu_{\beta} \). This together with (1) implies

\[
T' \vdash Con_{\tau'}.
\]

Thus we have \( Rfn(\tau) \equiv_T Rfn(\tau') \) as desired. \( \square \)
The following theorem shows that $\mathfrak{S}_T$ is a dense ordering of 
$\{Rfn(\tau) | \tau \in Bin(T, S)\}$.

3.3 Theorem. If $\tau_1, \tau_2 \in Bin(T, S)$ and $Rfn(\tau_1) \sqsubseteq_T Rfn(\tau_2)$, then there is a $\tau' \in Bin(T, S)$ for which 
$Rfn(\tau_1) \sqsubseteq_T Rfn(\tau') \sqsubseteq_T Rfn(\tau_2)$.

**Proof.** Since $Rfn(\tau_1) \sqsubseteq_T Rfn(\tau_2)$, there is a sentence $\phi \in \text{Sent}_T$ for which $Pr_{\tau_2}(\bar{\phi}) \rightarrow \phi$ is not provable in $T \cup Rfn(\tau_1)$. For this $\phi$, set $T' = T \cup Rfn(\tau_1) \cup \{\neg (Pr_{\tau_2}(\bar{\phi}) \rightarrow \phi)\}$. Then $T'$ is a consistent theory. Let $\beta'(x)$ be a $\Sigma_\kappa$-formula which binumerates $Rfn(\tau_1)$ in $PA$, and define $\beta(x)$ by 
$\beta(x) = \tau_1(x) \forall y \beta'(x) \forall y = \neg (Pr_{\tau_2}(\bar{\phi}) \rightarrow \phi).$

Clearly, $\beta$ is a binumeration of $T'$ in $S$. If we set

i) $\theta(x, y) = Prf_{\beta}(\bar{x}, y) \forall z \leq x \rightarrow Prf_{\beta}(y, z),$

ii) $\chi$: a fixed point of $\exists x \theta(x, y),$

iii) $\tau'(x) = \tau_1(x) \forall y \forall \exists y_1, y_2 < x (\theta(y_1, \bar{z}) \land Prf_{\tau_2}(\bar{\phi}, y_2) \land \neg \phi), then $\tau'$ is a bi-

numeration of $T$ in $S$. By Rosser's theorem,

$T \vdash \chi,$ \hspace{1cm} (1)

$T \not\vdash \neg \chi.$ \hspace{1cm} (2)

First we show that

$PA \vdash (Pr_{\tau_1}(\bar{\phi}) \rightarrow \phi) \land \neg (Pr_{\tau_2}(\bar{\phi}) \rightarrow \phi).$ \hspace{1cm} (3)

Note that

$PA \vdash \theta(y_1, \bar{z}) \land Prf_{\tau_2}(\bar{\phi}, y_2) \land \neg \phi \rightarrow \tau_1(y_2), y_2 < (\neg (Prf_{\tau_2}(\bar{\phi}, y_2) \land \neg (Prf_{\tau_2}(\bar{\phi}, y_2)))$

$\land \theta(y_1, \bar{z}) \land Prf_{\tau_2}(\bar{\phi}, y_2) \land \neg \phi$

$\rightarrow \neg \tau'(\neg (Prf_{\tau_2}(\bar{\phi}, y_2) \land \neg (Prf_{\tau_2}(\bar{\phi}, y_2)))$.

Then, clearly, we have

$PA \vdash \exists y_1 \theta(y_1, \bar{z}) \land Prf_{\tau_2}(\bar{\phi} \land \neg \phi) \rightarrow Pr_{\tau_1}(\bar{\phi}).$

This directly gives (3). Next we show that

$PA \vdash (Pr_{\tau_1}(\bar{\phi}) \rightarrow \phi) \land \neg \chi \rightarrow (Pr_{\tau_1}(\bar{\phi}) \rightarrow \phi)$ for all $\phi \in \text{Sent}_T.$ \hspace{1cm} (4)

But it is sufficient for this purpose to show that

$PA \vdash \neg \chi \rightarrow \forall x (\tau'(x) \rightarrow \tau_1(x)).$

And this is verified by the fact that $PA \vdash \neg \chi \rightarrow \forall x \neg \theta(x, \bar{z})$. Now, from (2) and (3), we have
On the other hand, from (1) and (4),

\[ T \cup Rfn(\tau_1) \vdash Pr_\tau(\bar{\phi}) \rightarrow \phi . \]  

Since \( Rfn(\tau_1) \subseteq T \cup Rfn(\tau_2) \) is clear from the equivalence of \( \tau' \) and \( \tau \), over 
\[ T \cup Rfn(\tau_2), \]  
(5) and (6) will complete our proof. \( \square \)

**Notes to Theorems 3.1, 3.2 and 3.3.** In the first two theorems, if \( \tau \) is a 
\( \Sigma^0 \)-binumeration of \( T \), we can choose \( \tau' \) in \( \Sigma^0 \). Hence there are also infinitely many kinds of the \( \tau \)-reflection principles, even if \( \tau \) is restricted to \( \Sigma^0 \). The author doesn't know whether \( \tau' \) of Theorem 3.3 can be chosen in \( \Sigma^0 \). To

The following theorem shows that there are incomparable elements in 
\[ \{ Rfn(\tau) | \tau \in Bin(T, S) \} \]  
with respect to \( \equiv_T \). Since its proof is very similar to that of 
Theorem 2.14 of [2], we shall give only a sketch of the proof.

3.4. **Theorem.** Suppose that \( T \) is an essentially reflexive and \( \omega \)-consistent theory. Then, for each \( \tau \in Bin(T, S) \), we can effectively construct a \( \tau' \in Bin(T, S) \) for which

\[ Rfn(\tau) \equiv_T Rfn(\tau') \]  
\& \[ Rfn(\tau') \equiv_T Rfn(\tau) . \]

**Sketch of the Proof.** Using Theorem 3.1, we choose \( \sigma \in Sent_T \) and \( \tau^* \in 
Bin(T, S) \) such that

\[ Rfn(\tau^*) \equiv_T Rfn(\tau) \]  
\& \[ T \cup Rfn(\tau^*) \vdash Pr_\tau(\bar{\sigma}) \rightarrow \sigma . \]

Putting \( A_1 = T \cup Rfn(\tau) \) and \( A_2 = T \cup Rfn(\tau^*) \cup \{ \neg (Pr_\tau(\bar{\sigma}) \rightarrow \sigma) \} \), we construct \( \tau_1 \in 
Bin(A_1, S) \) and \( \tau_2 \in Bin(A_2, S) \). Then we construct \( \tau' \) as

\[ \tau'(x) = \tau^*(x)^ \forall Fm(x, \mu) \exists y < x \neg M_{a_1, a_2}(\beta, y) , \]
where \( M_{a_1, a_2}(x, y) = (Pr_{a_1}(x, y) \lor Pr_{a_2}(x, y)) \rightarrow \exists z < y (Pr_{a_1}(x, z) \lor Pr_{a_2}(\neg x, z)) \)
and \( \mu \) is a fixed point of \( \forall y M_{a_1, a_2}(x, y) \). If we note that \( \mu \) is independent of 
\( A_1 \) and \( A_2 \), we can easily show the desired properties of \( \tau' \). \( \square \)

**Discussions.** The well-known theorem of Löb states that \( Pr_\tau(\bar{\phi}) \rightarrow \phi \) is provable in \( T \) if \( \phi \) is provable in \( T \). This does not contradict our results. The reason is that Löb's theorem holds only for r.e. theories \( T \) and their \( \Sigma^0 \)-numera-

\[ \text{tions} \ \tau. \]  
There are many results that hold only for r.e. theories, and it would be interesting to give examples of them. But we do not go further into these matters.
In this paper, we largely dealt with local reflection principles. The analogous results for uniform reflection principles will be contained in our forthcoming paper.

References


Department of Mathematics
University of Tsukuba
Sakura-mura, Niihari-gun
Ibaraki, 305 Japan