# An Unusual Reflection Principle for Self Justifying Logics

**Dan E. Willard**

University at Albany – SUNY

April 1, 2012

# 1. Overview

Gödel's 2nd Incomplet. Theorem indicates strong formalisms cannot verify their own consistency

*But Humans Intuitively Appreciate their Own Consistency*

Topic of our 64 Page Paper: What kinds of systems are Adequately Weak to formalize **some type (?)** of knowledge of their own consistency?

Research in New Technical Report and Six Prior Articles in JSL and APAL Has Sought to:

1. Develop New Generalizations of Second Inc Theorem
2. Formalize **Unusual** "Boundary-Case Exceptions" to It.
3. Produce Tightest Possible Match Between Items 1 + 2.

# 2. Background Literature (summarized in 3 slides)

Definition: Axiom System $\beta$ called **Self Justifying** relative to Deduction Method $d$ when :

1. one of $\beta$'s formal theorems states $d$'s deduction method, applied to axiom system $\beta$, is consistent.
2. and the axiom system $\beta$ is **also actually** consistent.

$\forall\ \alpha\ \ \forall\ d$ Kleene (1938), Rogers (1966) & Jersolow (1971) noted Easy To Construct axiom system $\alpha^d \supseteq \alpha$ satisfying Requirement 1

i.e. set $\alpha^d = \alpha \cup \text{SelfCons}(\alpha, d)$ (defined below)

"There is no proof (using $d$'s deduction method) of $0 = 1$ from the Union of system $\alpha$ with this sentence (looking at itself)"

Above Well Defined But Catch is $\alpha^d$ Usually Fails Item 2.

- i.e. $\alpha^d$ is inconsistent via a Gödel diagonalization paradigm.
- Thus prior to Willard (1993), this topic mostly shunned.

# 3. More Background Literature

Definition: Let $\alpha$ denote axiom system lacking Induction Principle
Then $\Psi(x)$ called $\alpha$-Initial Segment iff $\alpha$ can prove:

$$\Psi(0) \quad AND \quad \forall\, x \quad \Psi(x) \rightarrow \Psi(x+1) \tag{1}$$

Pudlák 1985: All axiom systems of finite cardinality have Initial Segments $\Psi$ where $\alpha$ can verify its Herbrand and Semantic Tableaux Consistency for every $x$ satisfying $\Psi(x)$

- Intuition: All integers $x$ satisfy $\Psi(x)$ BUT $\alpha$ NOT KNOW THIS !
- Above Result does not generalize for Hilbert Deduction

Kreisel-Takeuti (1974) Earliest Local-Consistency Result:

- Showed Second-Order Generalization of Cut-Free Deduction Can Verify Its Own Consistency.
- Sets $\Psi$ (in Equation 1) $=$ Dedekind's Definition of Integers

Verbrugge-Visser (1994) developed analogous arithmetic reflection principles using local consistency constructs.

- Visser (2005) discusses this topic further and summarizes Harvey Friedman's Ohio State 1979 Tech Report

# 4. Generalizations of Second Inc Theorem

- Bezboruah-Shepherdson 1976: Showed some Gödel encodings of Robinson's Q CANNOT VERIFY their Hilbert consistency.
- Pudlák 1985: Generalized Above for all Gödel encodings of proofs and for All Initial Segments (defined on prior slide) when Hilbert Deduction Present.
- Wilkie-Paris 1987 : showed $I\Sigma_0$+Exp CANNOT PROVE Hilbert Consistency of Q,
- Solovay (1994 Private Com.) : Showed **NO SYSTEM** (weaker than Q) Recognizing **MERELY SUCCESSOR** as total function can VERIFY its Hilbert Consistency.
- W— 2002-2009 : generalized work of Adamowicz-Zbierski to show THREE DIFFERENT ENCODINGS of $I\Sigma_0$ CANNOT PROVE their semantic tableaux consistency.

Hence Self-Justifying Formalisms **Always Contain** weaknesses.

# 5.Main Perspective of Willard's 1993-2009 Research

Notation: Add$(x, y, z)$ and Mult$(x, y, z)$ are 3-way atomic predicates employed by our axiom systems.

Definitions: An axiom system $\alpha$ is

- **Type-A** iff it contains Equation 1 as axiom:
- **Type-M** iff it contains 1 + 2 as axiom:
- **Type-S** iff it can prove (3) **BUT NOT PROVE** (1) NOR (2) :

$$\forall x \; \forall y \; \exists z \quad Add(x, y, z) \tag{1}$$

$$\forall x \; \forall y \; \exists z \quad Mult(x, y, z) \tag{2}$$

$$\forall x \;\; \exists z \quad Add(x, 1, z) \tag{3}$$

Combined Result of Pudlak, Solovay, Nelson, Wilkie-Paris:

- No natural Type-S system can recognize its Hilbert consistency:

Our Main Prior Results about this Subject:

1. Some Type-A prove all PA's $\pi_1$ theorems and their semantic tableaux consistency
2. Most Type-M axiom systems **UNABLE to JUSTIFY** their semantic tableaux consistency.

# 6. Limitations Upon Self Justifying Systems

1. Pudlak (1985) + Solovay (1994) (combined with Nelson + Wilkie-Paris) implies self-justication collapes when Hilbert Deduction is present for most systems rocognizing Successor as total functioon.
2. JSL(2002)+ APAL(2007) indicates Semantic Tableaux Self Jusitication collapses when Multiplication recognized as Total Function.
3. FOL-2004 Paper showed that while JSL 2005 could add a $\pi_1$ and $\Sigma_1$ modus ponens rule to our semantic tableaux evasions of Second Incompleteness Theorem, Same NOT TRUE with $\pi_2$ and $\Sigma_2$ modus ponens rules.

Next Three Slides Have GOOD NEWS despite Items 1-3:

*Self-Justifying Systems Support Unusually Robust Reflection Principles.*

Thus Bad News from Items 1-3 Not Fully Dismal !

# 7. New Perspective about Reflection Principles

Def: $\text{Reflect}_{\alpha,D}(\Psi)$ denotes sentence $\Psi$'s reflection principle under the axiom system $\alpha$ and deduction method $D$ i.e.

$$\forall\, p \quad \{ \quad \text{Prf}_{\alpha,D}(\ulcorner\Psi\urcorner\,,\; p\,) \quad \Rightarrow \quad \Psi \quad \} \qquad (4)$$

Löb's Theorem: If $\alpha \supset$ Peano Arith then $\alpha$ cannot prove $\text{Reflect}_{\alpha,D}(\Psi)$ except in trivial case where it can prove $\Psi$.

Gödel's Anti-Reflection Theorem: No reasonable axiom system $\alpha$ can prove $\text{Reflect}_{\alpha,D}(\Psi)$ for all $\pi_1$ sentences.

*i.e. Difficulties always arise because Gödel Sentences declaring "There is no proof of me" have $\pi_1$ encodings.*

Surprising Fact: Self-Justifying Systems Support "Transformed" $\pi_1$ Reflection Principles Despite Above 2 Theorems, i.e.

$$\forall\, p \quad \{ \quad \text{Prf}_{\alpha,D}(\ulcorner\Psi\urcorner\,,\; p\,) \quad \Rightarrow \quad \Psi^{T} \quad \} \qquad (5)$$

*where $T$ is isomorphism mapping $\pi_1$ sentences into $\pi_1$ sentences such that $\Psi \leftrightarrow \Psi^{T}$ holds in Standard Model.*

# 8. Two New Theorems About Reflection Principles

Def: Ax System $\alpha$ is Level( $1^D$ ) Consistent iff $\alpha$ UNABLE TO PROVE under deduction method $D$ BOTH some $\pi_1$ sentence and its negation.

Theorem 6.12 If $\alpha$ can formally verify its own Level( $1^D$ ) Consistency Then there exists some $T$ where $\alpha$ can verify (6)'s "Transformational" Reflection Principle for All $\pi_1$ sentences $\Psi$ simultaneously.

$$\forall\, p \quad \{ \ \mathrm{Prf}_{\alpha,D}( \ulcorner \Psi \urcorner , \ p \ ) \ \Rightarrow \ \Psi^T \ \} \tag{6}$$

Intuition Behind Theorem 6.12 : The identity $\Psi \leftrightarrow \Psi^T$ holds in Standard Model, BUT $\alpha$ UNABLE to verify it.

Theorem E.1 If Ax System $\alpha$ unable to prove its own consistency (i.e. satisfies Second Inc.Theorem) then $\alpha$ UNABLE TO VERIFY (6)'s Transform Reflection Principle for All $\pi_1$ sentences $\Psi$ simultaneously.

Proof Sketch: All conventional axiom systems can refute all false $\pi_1$ sentences. Hence if $\Psi$ false then $\alpha$ can refute both $\Psi$ and $\Psi^T$. But then $\alpha$ could use (6)'s reflection principle to confirm its own consistency. Latter impossible because contradicts Theorem 6.12's hypothesis. $\square$

# 9. Mysterious Two Sentences in Gödel's 1931 Paper

Most Surprising Two Sentences in Gödel's Paper:

- *"It must be expressly noted that Theorem XI (i.e the Second Inc Theorem) represents no contradiction of the formalistic standpoint of Hilbert. For this standpoint presupposes only the existence of a consistency proof by finite means, and there might conceivably be finite proofs which cannot be stated in ... "*

Our Interpretation of Gödel's Statement • :

1. We agree with most logicians that Gödel was excessively cautious in Statement • because history has proven the Second Inc Theorem to be a 95 % Robust Result from a "Consistency Perspective".
2. However, Gödel's Statement • is QUITE SIGNIFICANT from a "Reflection Perspective" because $\pi_1$ Transform Reflection explains how Thinking Beings aquire motivation to cogitate.

$$\forall\, p \quad \{ \;\; \text{Prf}_{\alpha,D}(\ulcorner \Psi \urcorner \,,\; p\;) \;\;\Rightarrow\;\; \Psi^{T} \;\; \} \tag{7}$$

# 10. Concluding Remarks

Wide Significance of Gödel's 2nd Incomp Theorem illustrated by:

- Its generalization using 1939 Hilbert-Bernays Derivation Conditions
- Solovay's 1994 Extension of Pudläk's 1985 Work:

  **No Axiom System** *viewing successor as a total function can justify its own Hilbert consistency.*

Above Precludes many but not all uses of **"I am consistent"** axioms:

1. This is because Reflection Principles explain how Thinking Beings Motivate Themselves to Cogitate
2. This use of Reflection Principles Is Very Helpful, EVEN IF it does not formalize a STRONG RESPECT where systems confirm their own consistency.

Many Other Results at http://arxiv.org/abs/1108.6330.

- Purpose of this Talk was to be pointer to 64-page report
- Latter Both Unifies and Extends our Prior Results