A computational lens on Self-Justifying Axiom Systems

Fellowship Bookends

Proposal:

> Within a host language, such as miniKanren, Lean, Idris, etc., encode the axiom basis and deductive apparatus of a specific SJASWithin a host language, such as miniKanren, Lean, Idris, etc., encode the axiom basis and deductive apparatus of a specific SJAS.

Deliverable:

A tableau style theorem prover for closed formulae of the $IS^{\lambda}(A)$ theory, written in miniKanren.

$IS^{\lambda}(A)$ ("isla")

A weak first order theory of arithmetic, with a semantic tableau deduction method.

Willard, D., 2001, *Self-Verifying Axiom Systems, the Incompleteness Theorem and Related Reflection Principles*

Subadditive Arithmetic

Integer Subtraction(x, y)
Integer Division(x, y)
Maximum(x, y)
Length(x) = $|x|$
Root(x, y) = $\lceil x\hat{}(1/y) \rceil$
Count(x, j) := the number of "1"s in x's rightmost j bits

Mult(x, y, z):=
$[(x = 0 \vee y = 0) \Rightarrow z = 0] \wedge$
$[(x \mathrel{!}= 0 \wedge y \mathrel{!}= 0) \Rightarrow (z/x = y \wedge ((z - 1)/x) < y)]$

Construct $\Delta_k, \Pi_k, \Sigma_k$ sentences as per normal.

Axiom Basis

Group 0: Formulae of first order subadditive arithmetic

Group 1: a finite set of Π_1 sentences, proving any Δ_0 sentence that holds true under the standard model

Group 2: For each Π_1 sentence Φ, $\forall p \{Prf(<\Phi>, p) \Rightarrow \Phi\}$

Group 3: Kleene Fixed Point encoding of consistency:

$$\forall y \neg SemPrf(<0=1>, y)$$

Tableau Deduction

Formula $f := \forall x \forall y \forall z\ (x + y \leq z) \Rightarrow (x \leq z \wedge y \leq z)$

$$
\begin{array}{ccccc}
 & & \neg f & & \\
/ & & | & \dots & \backslash \\
f_{11} & & f_{12} & & f_{1k} \\
| & & | & & | \\
\dots & & \dots & & \dots \\
| & & | & & | \\
f_{n1} \wedge \neg f_{n1} & & f_{n2} \wedge \neg f_{n2} & & f_{nk} \wedge \neg f_{nk}
\end{array}
$$

No modus ponens:

$a \Rightarrow b$
a
b

Self-Justification

1. Is consistent relative to an assumed consistent theory ("A", a parameter)

2. Can provably assert a statement of its own consistency:
   $\forall y\ \neg \mathrm{SemPrf}(<0=1>, y)$

Narrowly avoids Goedel's Second Incompleteness Theorem!

Motivation

Translating self-justification from the logical to the computational domain can improve our confidence in the correctness of software.

”Confidence in correctness” contributes to levelling the playing field in favor of the individual user and developer of software.

The Descent of Rigor

NTP Clock Strata:

Caesium-133 → Atomic Clock (0) → Time Server (1) → Local Server (2) → Personal Computer (3) → “about ten seconds”

Precision Measurement:

Molecular lattice → Hand-ground surface plate → Calibrated straightedge → Ruler → “it’s roughly level”

Software Correctness:

Mathematical Theory → Externally-audited proof kernel → Verified compiler → Source code → “looks good to me”

Software is Reflective

Mathematical Theory → Externally-audited proof kernel → Verified compiler → Source code → “looks good to me” → Mathematical Theory → ...

Human auditor: “I will not introduce inconsistencies”
– accepted or rejected according to informal methods

Machine auditor: “I will not introduce inconsistences”
– normally forbidden by G2.

Logical Limits

Goedel's Second Incompleteness Theorem (G2):

Any logical theory which is 1) sufficiently expressive, 2) effective, and 3) consistent is unable to 4) effectively demonstrate its own 5) consistency.

Logical Limits

Goedel's Second Incompleteness Theorem (G2):

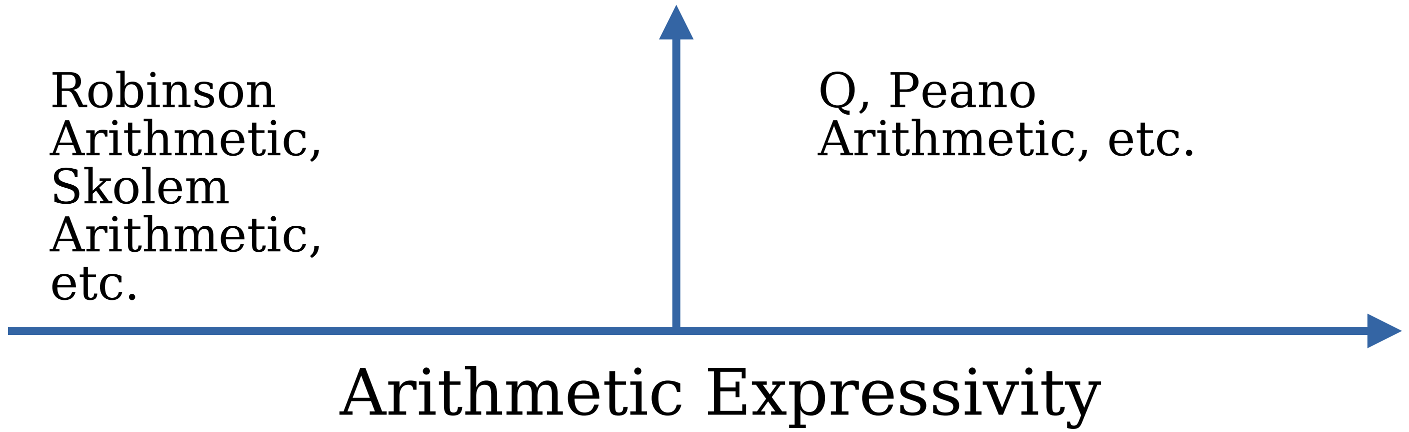
Any logical theory which is 1) sufficiently expressive, 2) effective, and 3) **consistent** is unable to 4) effectively demonstrate its own 5) consistency.

Logical Limits

Goedel's Second Incompleteness Theorem (G2):

Any logical theory which is 1) **sufficiently expressive**, 2) effective, and 3) **consistent** is unable to 4) effectively demonstrate its own 5) consistency.

Parametrizing Expressivity



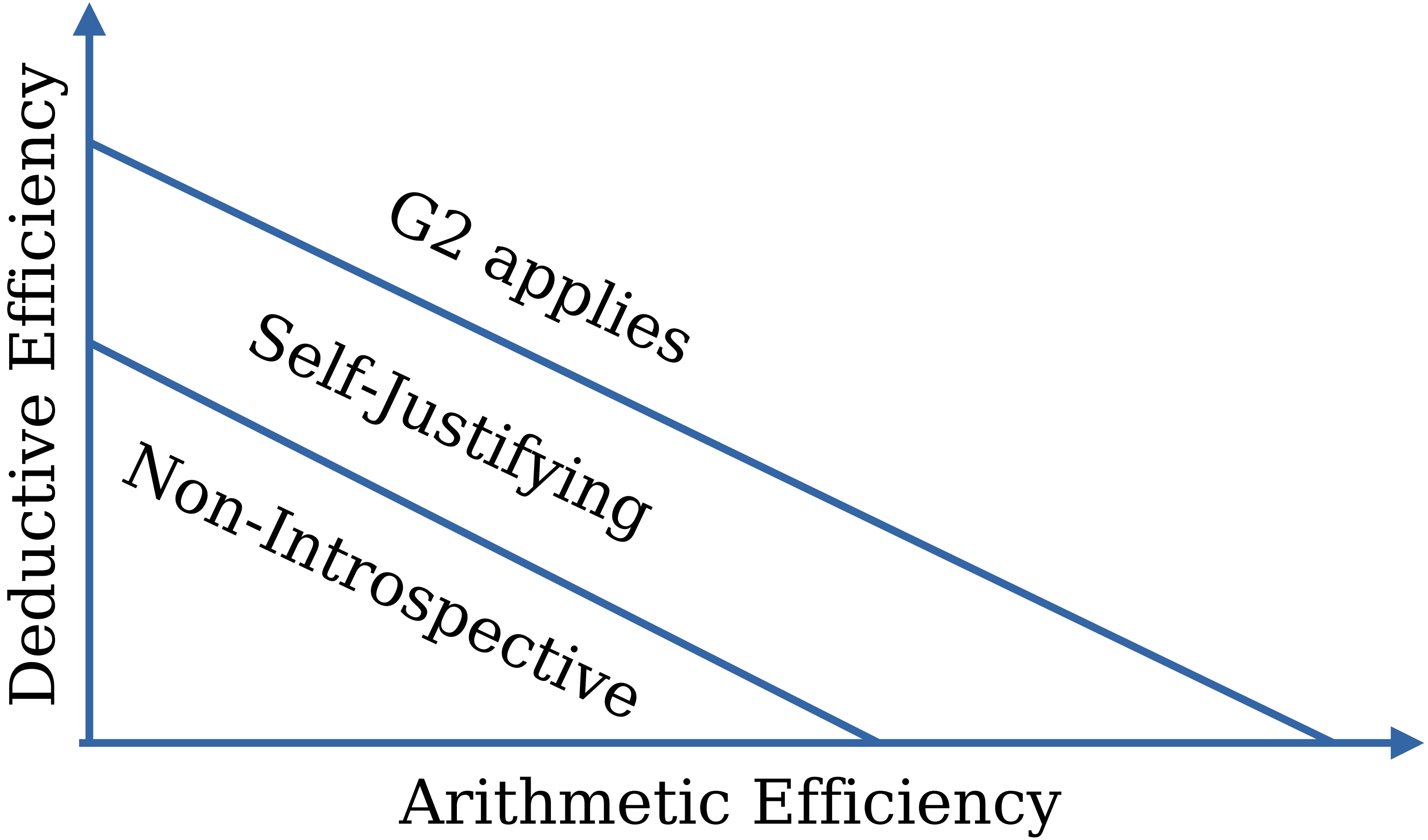
Parametrizing Expressivity

(1) $\forall x\, \exists z\, Add(x, 1, z)$
(2) $\forall x\, \forall y\, \exists z\, Add(x, y, z)$
(3) $\forall x\, \forall y\, \exists z\, Mult(x, y, z)$

NS: None
S: (1)
A: (1), (2)
M: (1), (2), (3)

”Linear” (Hilbert) vs “Tree style” (Tableau) deduction

Parametrizing Expressivity

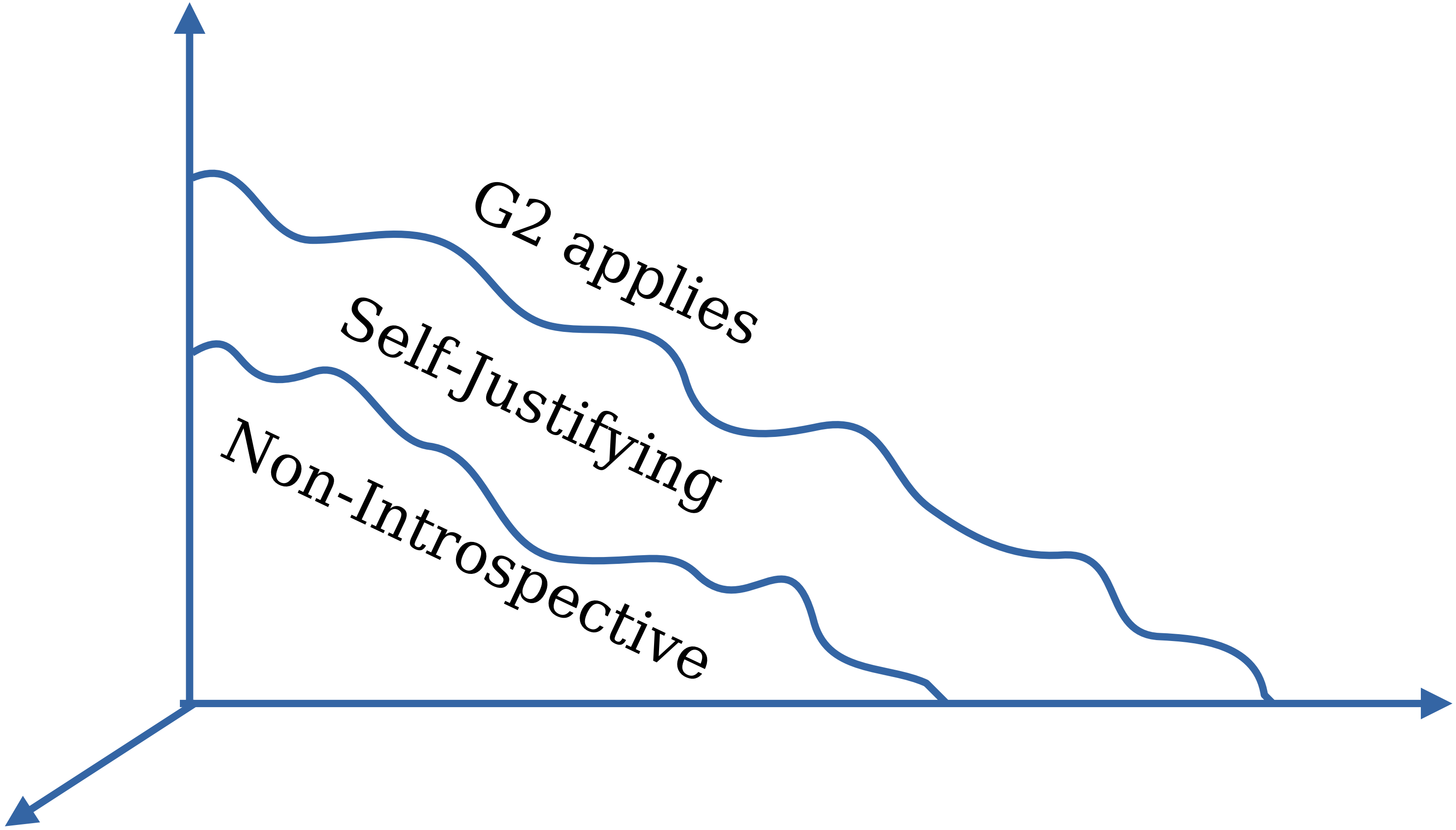


Fine-grained Parametrization

Name	Deduction Method	Type **A**	Almost **M**	Type **M**	**Axiom** Format	Self-Just Level
ξ^R	Resolution and/or Herbrandized analogs	Yes[35]	Yes	No	E-stable	Level (0^R)
ξ^*	Semantic Tableaux	Yes	No	No	EA-stable	Level (1^*)
ξ^{**}	Tab–U_1^* Deduction[34]	Yes	No	No	EA-stable	Level (1^*)
ξ^-	Hilbert Deduction	No	No	No	EA-stable	Level (∞^-)

Alternative Group-2 predicates?
Deduction methods using the cirquent calculus?

Fine-grained Parametrization



The Local Cluster

Artemov – proves each of an infinite tower of consistency statements for subsets of PA

Beklemishev – unpublished simplification of Willard’s SJAS

Ganea – SJAS does not define an intermediate r.e. degree

Niebergall – non-arithmetically definable, consistency proving theories

Pakhomov – “A weak set theory that proves its own consistency”

References

https://jpt401.substack.com/p/futo-fellowship-and-research-support

Willard, D., 2001, Self-Verifying Axiom Systems, the Incompleteness Theorem and Related Reflection Principles, http://www.jstor.org/stable/2695030

_, 2011, A Detailed Examination of Methods for Unifying, Simplifying and Extending Several Results About Self-Justifying Logics, https://arxiv.org/abs/1108.6330

_, 2005, An exploration of the partial respects in which an axiom system recognizing solely addition as a total function can verify its own consistency, DOI: 10.2178/jsl/1129642122

References

_, 2020, How the Law of Excluded Middle Pertains to the Second Incompleteness Theorem and its Boundary-Case Exceptions, https://arxiv.org/abs/2006.01057

Pakhomov, F., 2019, A weak set theory that proves its own consistency, https://arxiv.org/abs/1907.00877

Yanofsky, N. 2003, A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points, http://dx.doi.org/10.2178/bsl/1058448677

Additional information is available here: https://github.com/jpt4/sjas

Statement of Support

The isla theorem prover and related research was supported in part by the FUTO Fellows Summer 2025 program.

What and Why

Goedel's Second Incompleteness Theorem (G2):

Any logical theory which is 1) sufficiently expressive, 2) effective, and 3) consistent is unable to 4) effectively demonstrate its own 5) consistency.

What and Why

Goedel's Second Incompleteness Theorem (G2):

Any logical theory which is 1) sufficiently expressive, 2) effective, and 3) **consistent** is unable to 4) effectively demonstrate its own 5) consistency.

What and Why

Goedel's Second Incompleteness Theorem (G2):

Any logical theory which is 1) **sufficiently expressive**, 2) effective, and 3) **consistent** is unable to 4) effectively demonstrate its own 5) consistency.

What and Why

Self-Justifying Axiom System (SJAS):

i. A theory S includes a statement of the consistency of S.

ii. S is consistent relative to a known theory T.

Parametrizing Expressivity

Generalized Arithmetic Configuration:

i. Axiom Basis (language, consistency statement, et al.)

ii. Deductive Apparatus
- Logical Axioms
- Rules of Inference

Parametrizing Consistency

"An axiom system α owns a Level-1 appreciation of its own self- consistency (under a deductive apparatus D) iff it can verify that D produces no two simultaneous proofs for a Π_1 sentence and its negation."

Axiom Basis

First Order Language without Induction:

Integer Subtraction(x, y)
Integer Division(x, y)
Maximum(x, y)
Length(x) = |x|
Root(x, y) = ⌈x^(1/y)⌉
Count(x, j) := the number of “1”s in x’s rightmost j bits

Mult(x, y, z):=
$[(x = 0 \vee y = 0) \Rightarrow z = 0] \wedge [(x \mathrel{!=} 0 \wedge y \mathrel{!=} 0) \Rightarrow (z/x = y \wedge ((z - 1)/x) < y)]$

Construct Δ_k, Π_k, Σ_k sentences as per normal.

Axiom Basis

Group 0: Constants 0, 1, Addition(x, y), Double(x)

Group 1: a finite set of Π_1 sentences, proving any Δ_0 sentence that holds true under the standard model

Group 2: For each Π_1 sentence Φ, $\forall p \{\mathrm{HilbPrf}(<\Phi>, p) \Rightarrow \Phi\}$

Group 3: Kleene Fixed Point encoding of consistency:

$\forall x\ \forall y\ \forall p\ \forall q\ \neg[\mathrm{Pair}(x, y) \land \mathrm{Prf}(x, p) \land \mathrm{Prf}(y, q)]$

Deductive Apparatus

HilbPrf – Hilbert style proofs, using axiom schemae and modus ponens

Tab – standard semantic tableaux

X-Tab – semantic tableaux with an axiom scheme for LEM

Tab-1 – allows application of modus ponens for previously derived sentences.

Consistency Preservation

"An operation $I(\bullet)$ that maps an initial axiom system A onto an alternate system I(A) will be called Consistency Preserving iff I(A) is consistent whenever all of A's axioms hold true under the standard model of the natural numbers.

Suppose the symbol D denotes either semantic tableaux deduction or its Tab−1 generalization. Then the $I_D(\bullet)$ mapping operation is consistency preserving (e.g. $I_D(A)$ will be consistent whenever all of A's axioms hold true under the standard model of the natural numbers)."

Trade-Offs

(1) $\forall x\ \exists z\ Add(x, 1, z)$
(2) $\forall x\ \forall y\ \exists z\ Add(x, y, z)$
(3) $\forall x\ \forall y\ \exists z\ Mult(x, y, z)$

NS: None
S: (1)
A: (1), (2)
M: (1), (2), (3)

Tab-1 + A = Hilb + θ = SJAS

The Local Cluster

Artemov – proves each of an infinite tower of consistency statements for subsets of PA

Beklemishev – unpublished simplification of Willard’s SJAS

Ganea – SJAS does not define an intermediate r.e. degree

Niebergall – non-arithmetically definable, consistency proving theories

Pakhomov – “A weak set theory that proves its own consistency”

Further Questions

Is it correct?

Other parameter values

- Deduction systems: cirquent calculus, resolution

Executable realization

- Relational logic programming

Autarkic Formal Systems

To what extent are the aspects of a formal system determined by its own powers?

- Consistency
- Definability
- Decidability
- Interpretation
  Brown and Palsberg, self-interpretation in System F
- Replication
  Von Neumann automata

Autarkic Formal Systems

To what extent are the aspects of a formal system determined by its own powers?

- Consistency
- Definability <= Lawvere's Fixed Point Theorm (per Yanofsky)
- Decidability

Further motivation for intensional exploration.

Selected References

Willard, D., 2005, An exploration of the partial respects in which an axiom system recognizing solely addition as a total function can verify its own consistency, DOI: 10.2178/jsl/1129642122

_, 2020, How the Law of Excluded Middle Pertains to the Second Incompleteness Theorem and its Boundary-Case Exceptions, https://arxiv.org/abs/2006.01057

Pakhomov, F., 2019, A weak set theory that proves its own consistency, https://arxiv.org/abs/1907.00877

Yanofsky, N. 2003, A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points, http://dx.doi.org/10.2178/bsl/1058448677

Additional information is available here: https://github.com/jpt4/sjas