On the Significance of Self-Justifying Axiom Systems from the Perspective of Analytic Tableaux

Dan E.Willard[*]

State University of New York at Albany

Abstract

This article will be a continuation of our research into self-justifying systems. It will introduce several new theorems and their applications. (One of these results will transform our previous infinite-sized self-verifying formalisms into tighter systems, with only a finite number of axioms.) It will explain how self-justification is useful, even when the Incompleteness Theorem limits its reach.

Historical Remark about this Draft: All the theorems in this January 2014 manuscript had also appeared in the *"Version 1"* June 2013 manuscript that was posted earlier in the Cornell archives. The main difference between these two manuscripts is that I essentially spiced up the accompanying narrative, in the January 2014 draft, so as to better explain the motivation for this research endeavor.

Keywords: Gödel's Second Incompleteness Theorem, Consistency, Hilbert's Second Open Question, Semantic Tableaux

Mathematics Subject Classification: 03B52; 03F25; 03F45; 03H13

1 Introduction

Gödel's Incompleteness Theorem is a 2-part result. Its first half indicates no decision procedure can identify all the true statements of arithmetic. Its Second Incompleteness Theorem specifies sufficiently strong systems cannot verify their own consistency. Gödel was careful to insert the following caveat into his historic paper [13], indicating a diluted form of Hilbert's Consistency Program could be successful:

∗ : "It must be expressly noted that Proposition XI (e.g. the Second Incompleteness 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 P (or in M or in A). "

Yourgrau has summarized, in detail, Gödel's considerations about this subject. Thus [46] indicated that *"for several years"* after [13]'s publication, Gödel *"was cautious not to prejudge"* whether some unusual formalism, different from Peano Arithmetic, might provide some type of proof of its own consistency. Likewise, the Stanford Encyclopedia [29] cites Gödel remarking that it was only after the evolution of Turing's formalism [31] that Gödel viewed the Second Incompleteness Theorem as being fundamentally ubiquitous.

Within the framework of these particular caveats, our papers [34]–[45] have discussed both generalizations and boundary-case exceptions for the Second Incompleteness Theorem. This research has had two facets because the meaning of partial exceptions to the Second Incompleteness Theorem could be easily misunderstood, if their limitations are not also carefully scrutinized and recognized in meticulous detail.

The prior research was mostly mathematical in character, in that it sought to formalize several well-defined pairs (T, E), where T was a threshold sufficient for enacting the force of

the Second Incompleteness Theorem and E was a closely related "boundary-case exception" that omitted some part of T's formalism. Our goal in this paper will be different. It will seek to explore this subject from a more futuristic and epistemological perspective. It will characterize how far a theorem prover can traverse in understanding its own consistency before it reaches the inevetible barriers imposed by the Incompleteness Theorem. It will provide some partially sympathetic interpretations of Hilbert [16]'s 1925 statement ∗∗, in a context where Gödel [13] established it, obviously, needed serious amendments:

> ∗∗ "Where else would reliability and truth be found if even mathematical think-
> ing fails? The definitive nature of the infinite has become necessary, not merely
> for the special interests of individual sciences, but rather for the honor of human
> understanding itself."

In a context where Gödel's statement ∗ was *partially* sympathetic with Hilbert's goals, *even after* Gödel had proven his incompleteness theorem, we will show how some logic systems can support a positive *but-curtailed* form of Hilbert's objectives. From both a theoretical and a pragmatic engineering-style perspective, our results will helpfully explain how to transform [34, 36, 41, 42, 44, 45]'s results about infinite-sized self-justifying systems into compact tableaux formalisms, containing only a *strictly finite* number of proper axioms.

These results will suggest that although the Second Incompleteness Theorem is extraordinarily robust and ubiquitous from an idealized purist mathematical perspective, some unusual axiom systems can possess a fragmentary knowledge of their consistency, when using definitions of consistency that are diluted *but not immaterial.*

This research was partially influenced by Goldstein's biography of Gödel [14]. It suggested that there ought to be some type of partial philosophical compromise available between the mathematical Platonism of Gödel and the logical positivism of Wittgenstein.

Also, the style of presentation in this article was influenced by a suggestion that Selmer Bringsjord had made several year ago [4]. It was that our research should be ideally broken into two stages, with its first phase focusing on purely mathematical results (similar to [34]–[45]'s treatment), and with its second stage exploring the epistemological and futuristic implications of such results. Bringsjord felt this carefully segmentized approach was useful so that the mathematical novelty of our results not be confused with its philosophical interpretation. (This is because the latter, quite naturally, lends itself much more easily to a quite complicated debate about the underlying epistemological implications of our mathematical formalism.)

In essence, the current paper will bring this second stage of our research to completion. It will explain the significance of logics that formalize a partial, *although admittedly fragmentary,* knowledge of their own consistency.

Format of our Presentation: As much as is feasible, this article will be written in an informal style, so as to make it comprehensible to a broader audience. A reader who has, thus, mastered either say Enderton's or Mendelson's introductory logic textbooks [8, 23], during a 1-semester course in logic, should be able follow its gist.

Section 11 will suggest that our fragmentized approach for interpreting the goals of an applications-oriented engineering-styled subset of mathematics be called "Miniaturized Finitism". The cautious-sounding adjective of *"miniaturized"* was attached to our formalism's name partly because it involves a curtailed notion of growth (e.g. see Sections 7 and 8) and also because the classic variants of the Second Incompleteness Theorem, obviously, have resolved at least 90% of the issues raised by Hilbert's Second Open Question.

Our discourse will focus, essentially, on the remaining, say, 10 % of the issues raised by Hilbert's year-1900 penetrating Second Open Question. Thus, we seek to offer a novel interpretation of Gödel's and Hilbert's 1925 and 1931 statements of $*$ and $**$, that will explain how human beings manage to gain the motivation and mental energy required for stimulating their cogitations, by having a type of automatic and psychologically unconscious access to an

essentially *very very* miniaturized but-spontaneous self-appreciation of their own consistency.

2   Background Setting

Throughout this article, α will denote an axiom system, and d will denote a deduction method. An ordered pair (α, d) will be called **Self Justifying** when:

i   one of α's theorems will state that the deduction method d, applied to the system α, will produce a consistent set of theorems, and

ii   the axiom system α is in fact consistent.

For any (α, d), it is easy to construct a second axiom system $\alpha^d \supseteq \alpha$ that satisfies Part-i of this definition. For instance, α^d could consist of all of α's axioms plus the following added sentence, that we call **SelfRef**(α, d) :

- There is no proof (using d's deduction method) of $0 = 1$ from the *union* of the axiom system α with *this* sentence "SelfRef(α, d)" (looking at itself).

Kleene [19] discussed how to encode approximate analogs of SelfRef(α, d)'s self-referential statement. Each of Kleene, Rogers and Jeroslow [19, 27, 18] noted α^d may, however, be inconsistent (despite SelfRef(α, d)'s assertion), thus causing it to violate Part-ii of self-justification's definition.

This problem arises in settings more general than Gödel's paradigm, where α was an extension of Peano Arithmetic. There are many settings where the Second Incompleteness Theorem does generalize [1, 2, 3, 5, 6, 13, 15, 17, 21, 22, 25, 26, 27, 28, 30, 32, 33, 37, 39, 42, 43]. Each such result formalizes a paradigm where self-justification is infeasible, due to a diagonalization issue. Many logicians have, thus, hesitated to employ a SelfRef(α, d) axiom because α+SelfRef(α, d)

4

is typically inconsistent [1].

Our research explored special circumstances [36, 40, 41, 42] where it is feasible to construct self-justifying formalisms. These paradigms involved weakening the properties a system can prove about addition and/or multiplication (to avoid the preceding difficulties). To be more precise, let $Add(x, y, z)$ and $Mult(x, y, z)$ denote two 3-way predicates specifying $x + y = z$ and $x * y = z$. Then a logic will be said to **recognize** successor, addition and multiplication as **Total Functions** iff it includes 1-3 as axioms.

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

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

$$\forall x \; \forall y \; \exists z \quad Mult(x, y, z) \tag{3}$$

A logic system α will be called **Type-M** iff it contains $(1) - (3)$ as axioms, **Type-A** iff it contains only (1) and (2) as axioms, and **Type-S** iff it contains only (1) as an axiom. Also a system is called **Type-NS** iff it contains none of these axioms. The significance of these constructs is explained by items (a) and (b):

a. The existence of Type-A systems that can recognize their consistency under semantic tableaux deduction was proven in [40]. Also, [42] demonstrated a large class of Type-NS systems can recognize their Hilbert consistency. (Many of these systems do prove all Peano Arithmetic's Π_1 theorems in a language that represents addition and multiplication as 3-way predicates.)

b. The above evasions of the Second Incompleteness Theorem are known to be near-maximal in a mathematical sense. This is because the combined work of Pudlák, Solovay, Nelson and

---

[1] Typically, $\alpha^d = \alpha + \text{SelfRef}(\alpha, d)$ will be inconsistent because a standard Gödel-like self-referencing construction will produce a proof of $0 = 1$ from α^d, *even when* α *is consistent.*

Wilkie-Paris [24, 26, 28, 33] implied no natural Type-S system can recognize its Hilbert consistency, and Willard [37, 43, 44] strengthened some earlier results by Adamowicz-Zbierski [1, 2] to establish that most Type-M systems cannot recognize their semantic tableaux consistency.

An unusual aspect of (a) and (b) is that there is a tight match between their positive and negative results from a purely quantitative perspective, but they still do not reach one's ideal goals for deduction. This is because there is something about the deep yearnings of research, roughly suggested by Hilbert in ∗∗ and Gödel in ∗, which has not been addressed. This yearning and its tiny remaining gap will be our main focus in this article.

Other efforts to evade the Second Incompleteness Theorem have used the Kreisel-Takeuti "CFA" system [21] or what could be called the interpretational framework of Friedman, Nelson, Pudlák and Visser [12, 24, 26, 32]. These methods do not use Kleene-like *"I am consistent"* axioms, similar to those in our work. Instead, CFA uses the unique properties of Second Order generalizations of Gentzen's Sequent Calculus (with modus ponens absent), and interpretational frameworks formalize how some systems recognize their Herbrand consistency on localized sets of integers (which unbeknownst to themselves) includes all natural numbers.

Such systems are not germane to our exploration of Kleene-like *"I am consistent"* axioms, in the current article, but they do illustrate alternative approaches that are germane to other fascinating open questions about Incompleteness paradigms.

As the reader exams the next several sections of this paper, he should also keep in mind that self-justifying systems that rely upon semantic tableaux as their primary deduction method is a different topic than formalisms using Herbrand-styled deduction. This fact was brought to our attention by private communications from L. A. Kołodziejczyk [20], who noted that Herbrand styled systems can be exponentially slower than semantic tableaux systems under well-defined circumstances. Kołodziejczyk's insightful observation enabled us to prove in [44] that Herbrand-styled self-justifying systems, unlike their semantic tableaux counterparts, could

house a multiplication function symbol. Our results about mutliplication in [44] are clearly of mathematical interest, but they are not germane to our philosophical orientation in the current article because [44]'s formalism is able to house a multiplication function symbol only because of its exponential increase in inefficency. Thus, the current article will focus instead on uses of [40, 41]'s semantic tableaux styled formalisms because we suspect that their added efficiencies are likely to be of greater significance, from both a philosophical and engineering perspective, than our slower and more inefficient alternative mechanisms in [44].

3 Defining Notation

The next two sections will summarize the properties of [40]'s $\text{IS}_D(A)$ axiom system. We will then outline how to refine [40]'s results in the remainder of this article.

A function $F(a_1, a_2...a_j)$ will be called **Non-Growth** iff it satisfies the general inequality of $F(a_1, a_2, ...a_j) \leq Maximum(a_1, a_2, ...a_j)$. Six examples of non-growth functions are *Integer Subtraction* (where $x - y$ is defined to equal zero when $x \leq y$), *Integer Division* (where $x \div y$ is defined to equal x when $y = 0$, and it equals $\lfloor x/y \rfloor$ otherwise), $Maximum(x, y)$, $Logarithm(x)$, $Root(x, y) = \lceil x^{1/y} \rceil$ and $Count(x, j)$ designating the number of "1" bits among x's rightmost j bits.

The term **U-Grounding Function** referred in [40] to a set of eight operations, which included the preceding functions plus the *growth operations* of addition and $Double(x) = x + x$. Our language L^* was built out of these function symbols, the usual predicates of "$=$" and "$\leq$" and the constant symbols "0" and "1".

In a context where t is any term in [40]'s language L^*, the quantifiers stored in the wffs of $\forall v \leq t \ \Psi(v)$ and $\exists v \leq t \ \Psi(v)$ were called *bounded quantifiers*. Also, any formula in the U-Grounding language, all of whose quantifiers are bounded, was called a Δ_0^* formula. The Π_n^* and Σ_n^* formulae were then defined by the usual rules that:

1. Every Δ_0^* formula is considered to be a " Π_0^* " and also to be a " Σ_0^* " formula.

2. For $n \geq 1$, a formula is called Π_n^* when it is encoded as $\forall v_1 \ldots \forall v_k \; \Phi$ with Φ being Σ_{n-1}^*

3. Likewise, a formula is called Σ_n^* when it is encoded as $\forall v_1 \ldots \forall v_k \; \Phi$, with Φ being Π_{n-1}^*.

Example 3.1 Although our language L^* contains no multiplication function symbol, it does provide a means to encode multiplication as a 3-way relation Δ_0^* formula, $\mathrm{Mult}(x, y, z)$, as is illustrated below:

$$[\, (x = 0 \vee y = 0) \Rightarrow z = 0 \,] \;\; \wedge \;\; [\, (x \neq 0 \wedge y \neq 0 \,) \;\Rightarrow\; (\, \frac{z}{x} = y \;\wedge\; \frac{z-1}{x} < y \;) \,] \tag{4}$$

Moreover, (5) illustrates how the commutative principle for multiplication can receive a Π_1^* encoding via the above $\mathrm{Mult}(x, y, z)$ predicate.

$$\forall \, x \;\; \forall \, y \;\; \forall \, z \quad \mathrm{Mult}(x, y, z) \;\Leftrightarrow\; \mathrm{Mult}(y, x, z) \tag{5}$$

Similar methodologies may also encode multiplication's associative and distributive axioms as Π_1^* sentences. In general, all the Π_1 theorems in a conventional arithmetic language (that has multiplication symbols) can be translated into equivalent Π_1^* statements in our language L^*. This implies that the set of Π_1^* sentences is a quite rich class of statements and that no decision procedure is capable of separating all true from false Π_1^* statements.

Further Terminology: Most of our other notation will be similar to [40]'s terminology. In addition to considering a definition of semantic tableaux deduction that is similar to either Fitting's classic formalism [9] or a variant of it in [40], we considered in [40] also a somewhat stronger technique, called $\mathrm{Tab}-k$ deduction. It is defined below and consists of a speeded-up version of a tableaux proof, which permits an analog of modus ponens for the limited cases of performing Gentzen-style deductive cuts on Π_k^* and Σ_k^* formulae.

Definition 3.2 Let H denotes a sequence of ordered pairs (t_1, p_1), (t_2, p_2), ... (t_n, p_n), where p_i is a Semantic Tableaux proof of the theorem t_i. Then H will be called a **Tab-k Proof** of a theorem T from the axiom system α iff $T = t_n$ and also:

1. Each of the "intermediately derived theorems" $t_1, t_2, \ldots, t_{n-1}$ must have a complexity no greater than that of either a Π_k^* or Σ_k^* sentence.

2. Each axiom in p_i's proof either comes from α or is one of $t_1, t_2, \ldots, t_{i-1}$.

Thus, the preceding 2-part definition implies a Tab$-k$ proof differs from Fitting's definition of a semantic tableaux proof [9] by allowing for an analog of a Gentzen cut-rule to be applied to intermediate results at the levels of Π_k^* and Σ_k^* formulae.

Remark 3.3 Let us say that an axiom system α has a **Level-J Understanding** of its own consistency under a particular deduction method D iff α can prove that there exists no proofs using its axioms and D's deductive methodology of both a Π_J^* theorem and its negation. In this notation, items A and B summarize [35, 37, 39, 40, 43]'s results:

A. For every axiom system A using L^*'s U-Grounding language, [40] showed its $\text{IS}_D(A)$ formalism could prove all A's Π_1^* theorems and simultaneously verify its own Level-1 consistency when D corresponds to Tab-1 deduction.

B. Two negative results that tightly complemented item A's positive result were exhibited in [35, 37, 39, 43]. The first was that [35, 37, 43] showed most systems are unable to verify their Level-0 consistency under semantic tableaux deduction when they included statement (3)'s "Type-M" axiom that multiplication is a total function. Moreover, [39] offered an alternate form of the Second Incompleteness Theorem that showed statement (2)'s *far weaker* Type-A systems are unable to verify their Level-0 consistency under Tab-2 deduction.

The contrast between these positive and negative results has led to our conjecture that sophisticated theorem provers are likely to eventually achieve a fragmentary part of the ambitions stated by Gödel and Hilbert in $*$ and $**$. This is because the question of whether a formalism can support an *idealized Utopian* conception of its own consistency is *different* from exploring

the degrees to which theorem-provers can possess a *fragmentary knowledge* of their own consistency. Thus, the Incompleteness Theorem has demonstrated an Utopian idealized form of self-justification is unobtainable, but our on-going research has found some restrictive forms of self-knowledge are, indeed, feasible.

4   The $\mathrm{IS}_D(A)$ Axiom System

In a context where D denotes any deductive method and A denotes any axiom system using L^*'s U-Grounding language, $\mathrm{IS}_D(A)$ was defined in [40] to be an axiomatic formalism capable of recognizing all of A's Π_1^* theorems and corroborating its own Level-1 consistency under deductive method D. It was defined in [40] to consist of the following four groups of axioms:

Group-Zero: Two of the Group-zero axioms will be Π_1^* statements defining the named constant-symbols, $\bar{c}_0$ and $\bar{c}_1$, that designate the integers of 0 and 1. The third and fourth Group-zero axioms will be Π_1^* statements defining the two standard growth functions of addition and $Double(x) = x + x$. The net effect of these axioms will be to set up a machinery to define any integer $n \geq 2$ using fewer than $3 \cdot \lceil \mathrm{Log}\ n \rceil$ logic symbols.

Group-1: This axiom group will consist of a finite set of Π_1^* sentences, denoted as F, which which can prove any Δ_0^* sentence that holds true under the standard model of the natural numbers. (Any finite set of Π_1^* sentences F with this property may be used to define Group-1, as [40] noted.)

Group-2: Let $\ulcorner \Phi \urcorner$ denote Φ's Gödel Number, and $\mathrm{HilbPrf}_A(\ulcorner \Phi \urcorner, p)$ denote a Δ_0^* formula indicating p is a Hilbert-styled proof of theorem Φ from axiom system A. For each Π_1^* sentence Φ, the Group-2 schema will contain an axiom of form (6). (Thus $\mathrm{IS}_D(A)$ can trivially prove all A's Π_1^* theorems.)

$$\forall\ p\ \ \{\ \mathrm{HilbPrf}_A(\ulcorner \Phi \urcorner, p)\ \ \Rightarrow\ \ \Phi\ \} \tag{6}$$

Group-3: The final part of the $\mathrm{IS}_D(A)$ will consist of a single self-referencing Π_1^* sentence, indicating $\mathrm{IS}_D(A)$ meets §3's criteria of being "Level-1 consistent" under deductive method D. It is, thus, the following declaration:

> \# There exists no two proof of both a Π_1^* sentence and its negation when D's deductive method is applied to an axiom system that consists of the *union* of the Group-0, 1 and 2 axioms with *this sentence looking at itself.*

One encoding of \#, as a self-referencing Π_1^* sentence, was provided in [40]. Thus, sentence (7) will be a Π_1^* encoding for \#, in a context where $\mathrm{Prf}_{\mathrm{IS}_{D(A)}}(a, b)$ is a Δ_0^* formula indicating that b is a proof of a theorem a via deduction method D from $\mathrm{IS}_D(A)$'s axiom system, and where $\mathrm{Pair}(x, y)$ is a Δ_0^* formula indicating x is the Gödel number of a Π_1^* sentence and y represents x's negation.

$$\forall\, x\, \forall\, y\, \forall\, p\, \forall\, q \quad \neg\, [\ \mathrm{Pair}(x, y)\ \wedge\ \mathrm{Prf}_{\mathrm{IS}_{D(A)}}(x, p)\ \wedge\ \mathrm{Prf}_{\mathrm{IS}_{D(A)}}(y, q)\ ] \qquad (7)$$

Notation. 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. In this context, [40] demonstrated:

Theorem 4.1 *Suppose the symbol D denotes either semantic tableaux deduction or its Tab-1 generalization (given in Definition 3.2). Then the $\mathrm{IS}_D(\ \bullet\ )$ mapping operation is consistency preserving (e.g. $\mathrm{IS}_D(A)$ will be consistent whenever all of A's axioms hold true under the standard model of the natural numbers).*

We emphasize that the most difficult part of [40]'s result was neither the definition of its $\mathrm{IS}_D(A)$'s axiom system nor the Π_1^* fixed-point encoding of (7)'s Group-3 axiom. Instead, it was the confirming of Theorem 4.1 "Consistency Preservation" property.

The confirming of this property was subtle because its invariant breaks down when D is a deduction method only slightly stronger than either semantic tableaux or Definition 3.2's Tab-1 construct. Thus, the combined work of Pudlák and Solovay [26, 28] implied that Theorem 4.1's analog fails when D represents Hilbert deduction, and [39] showed its generalization would even fail when D represents Tab-2 deduction.

These difficulties occur because a Gödel-like diagonalization effect will automatically render $\text{IS}_D(A)$ inconsistent when D is too strong. Throughout this article, we will need to assure D does not reach a strength that produces such issues.

5   A Finitized Generalization of Theorem 4.1's Methodology

One awkward aspect of $\text{IS}_D(A)$ is that it employs an infinite number of different incarnations of sentence (6)'s Group-2 schema (since it contains one incarnation of this sentence for each Π_1^* sentence Φ in L^*'s language). Such a Group-2 schema is cumbersome because it simulates A's Π_1^* knowledge almost via a brute-force enumeration.

Our Definition 5.1 and Theorems 5.2 and 5.7 will show how to mostly overcome this difficulty by compressing the infinite number of instances of sentence (6) in $\text{IS}_D(A)$'s Group-2 schema into a purely finite structure.

Definition 5.1 Let β denote any finite set of axioms that have Π_1^* encodings. Then $\text{IS}_D^{\#}(\beta)$ will denote an axiom system, similar to $\text{IS}_D(A)$, except its Group-2 scheme will employ β's set of axioms, instead of using an infinite number of applications of statement (6)'s scheme. (Thus, the *"I am consistent"* statement in $\text{IS}_D^{\#}(\beta)$'s Group-3 axiom will be the same as before, except that the *"I am"* fragment of its self-referencing statement will reflect these changes in Group-2 in the obvious manner.)

Our next theorem will indicate that an analog of Theorem 4.1's consistency preservation property applies to $\text{IS}_D^{\#}(\beta)$'s formalism:

Theorem 5.2 *Let D again denote either semantic tableaux deduction or Definition 3.2's Tab-1 construct, and β again denote a set of Π_1^* axioms. Then $IS_D^\#(\beta)$ will be consistent whenever all β's axioms hold true under the standard model. (In other words, $IS_D^\#(\beta)$ will satisfy an analog of Theorem 4.1's consistency preservation property for $IS_D(A)$.)*

A formal proof of Theorem 5.2 from first principles is essentially as lengthy as [40]'s proof of Theorem 4.1. It is unnecessary to provide such a proof here because both theorems are justified using roughly analogous methods. Appendix A offers a brief summary about how [40]'s proof of Theorem 4.1 can be easily incrementally modified to also prove Theorem 5.2.

Our goal in the current article will be to show how Theorem 5.2's $IS_D^\#(\beta)$ formalism can assure that self-justifying axiom systems can prove an astonishingly wide breadth of theorems, *even when such logics* contain only a *strictly finite* number of axiomatic statements. We need one further definition to explore the depth of this meta-result:

Definition 5.3 Let $\ulcorner \Psi \urcorner$ denote Ψ's Gödel number. A Δ_0^* formula, $\text{Test}_i(t, x)$, will be called a **Kernelized Formula** iff Peano Arithmetic can prove every Π_1^* sentence Ψ satisfies (8)'s identity:

$$\Psi \iff \forall x \ \text{Test}_i(\ulcorner \Psi \urcorner, x) \tag{8}$$

There are countably infinitely many different Δ_0^* predicates $\text{Test}_1(t, x)$, $\text{Test}_2(t, x)$, $\text{Test}_3(t, x)$ ... that satisfy this kernelized condition. (One such kernel predicate is illustrated by Example 5.4 below). A formal enumerated list of all such kernel predicates will be called a **Kernel-List**.

Example 5.4 The set of true Σ_1^* sentences is clearly recursively enumerable. This easily implies there exists a Δ_0^* formula, called say $\text{Probe}(g, x)$, such that g is the Gödel number of a Σ_1^* statement that holds true in the Standard Model if and only if (9) is true:

$$\exists x \ \ \text{Probe}(g, x) \ \wedge \ x \geq g \tag{9}$$

Now, let $\text{Pair}(t, g)$ denote a Δ_0^* formula that specifies t is the Gödel number of a Π_1^* statement and g is the Σ_1^* formula which is its negation. Then our notation implies that t is a true Π_1^*

statement if and only if (10) holds true:

$$\forall\, x \quad \neg\, [\, \exists\, g \,\leq\, x \qquad \mathrm{Pair}(t,g) \,\wedge\, \mathrm{Probe}(g,x) \,\,] \tag{10}$$

Thus if $\mathrm{Test}_0(t,x)$ denotes the Δ_0^* formula of " $\neg\, [\, \exists\, g \,\leq\, x\ \ \mathrm{Pair}(t,g) \,\wedge\, \mathrm{Probe}(g,x)]$ ", then it is one example of what Definition 5.3 would call a "Kernelized Formula".

Definition 5.5 Let us recall that Definition 5.3 defined **Kernel-List** to be an enumeration of all the available kernelized formulae of $\mathrm{Test}_1(t,x)$, $\mathrm{Test}_2(t,x)$, $\mathrm{Test}_3(t,x)$ ... Then if $\mathrm{Test}_i(t,x)$ is the $i-$th element in this list of kernels and if Ψ is an arbitrary Π_1^* sentence, then Ψ's **i-th Kernel Image** will be the following Π_1^* sentence:

$$\forall\, x \ \ \mathrm{Test}_i(\,\ulcorner \Psi \urcorner\,,\, x\,) \tag{11}$$

Example 5.6 The combination of Definitions 5.3 and 5.5 indicates that there is a subtle relationship between a sentence Ψ and its $i-$th Kernel Image. This is because Definition 5.3 indicates that Peano Arithmetic does prove the invariant (8), indicating that Ψ is equivalent to its its $i-$th Kernel Image. However, a weak axiom system can be plausibly uncertain about whether this equivalence holds.

Thus if a weak axiom system proves statement (11) (rather than Ψ itself), then it may not be able to equate these two results. This problem will apply to Theorem 5.7's formalism. However, Theorem 5.7 will be still of much interest because §6 will illustrate a generalized methodology that overcomes many of Theorem 5.7's limitations.

Theorem 5.7 *Let A denote any system, all of whose axioms hold true in arithmetic's standard model, and i denote the index of any of Definition 5.3's kernelized formulae $\mathrm{Test}_i(t,x)$. Then it is possible to construct a finite-sized collection of Π_1^* sentences, called say $\beta_{A,i}$, where $\mathrm{IS}_D^\#(\beta_{A,i})$ satisfies the following invariant:*

If Ψ is one of the Π_1^* theorems of A then $\mathrm{IS}_D^\#(\beta_{A,i})$ can prove (11)'s statement

(e.g. it will prove the "the $i-$th kernelized image" of the sentence Ψ).

Proof Sketch: Our justification of Theorem 5.7 will use the following notation:

1. Check(t) will denote a Δ_0^* formula that produces a Boolean value of True when t represents the Gödel number of a Π_1^* sentence.

2. HilbPrf$_A$ (t, q) will denote a Δ_0^* formula that indicates q is a Hilbert-style proof of the theorem t from axiom system A.

3. For any kernelized Test$_i(t, x)$ formula, the symbol GlobSim$_i$ will then denote (12)'s Π_1^* sentence. (It will be called A's $i-$th **"Global Simulation Sentence"**.)

$$\forall\, t \;\; \forall\, q \;\; \forall\, x \;\; \{ \;\; [\;\; \text{HilbPrf}_A\,(t,q) \;\; \wedge \;\; \text{Check}(t) \;\;] \;\; \implies \;\; \text{Test}_i(t,x) \;\; \} \tag{12}$$

In this notation, the requirements of Theorem 5.7 will be satisfied by any version of the axiom system $\text{IS}_D^{\#}(\beta)$, whose Group-2 schema β is a finite sized consistent set of Π_1^* sentences that has (12) as an axiom. (This includes the minimal sized such system, that has only (12) as an axiom.) This is because if Ψ is any Π_1^* theorem of A, whose proof is denoted as $\bar{p}$, then both the Δ_0^* predicates of HilbPrf$_A$ ($\ulcorner\Psi\urcorner, \bar{p}$) and Check($\ulcorner\Psi\urcorner$) are true. Moreover, $\text{IS}_D^{\#}(\beta)$'s Group-1 axiom subgroup was defined so that it can automatically prove all Δ_0^* sentences that are true. Thus, $\text{IS}_D^{\#}(\beta)$ will prove these two statements and hence corroborate (via axiom (12)) the further statement:

$$\forall\, x \;\; \text{Test}_i\,(\; \ulcorner\Psi\urcorner \;,\; x\;) \tag{13}$$

Hence for each of the infinite number of Π_1^* theorems that A proves, the above defined formalism will prove a matching statement that corresponds to the $i-$th kernelized image of each such proven theorem. $\square$

6   Pragmatic L-Fold Generalizations of Theorem 5.7

Theorem 5.7 is of interest because every axiom system A will have its formalism $\text{IS}_D^{\#}(\beta_{A,i})$ prove the $i-$th kernelized image of every Π_1^* theorem that A proves. This fact is helpful because

(8)'s invariance holds for all Π_1^* sentences. Moreover, our "U-Grounded" Π_1^* sentences capture all Conventional Arithmetic's *crucial* Π_1 information because they can view multiplication as a 3-way Δ_0^* predicate $\text{Mult}(x, y, z)$ via (14)'s encoding of this predicate.

$$[\, (x = 0 \vee y = 0) \Rightarrow z = 0 \,] \ \wedge \ [\, (x \neq 0 \wedge y \neq 0 \,) \ \Rightarrow \ (\, \frac{z}{x} = y \ \wedge \ \frac{z-1}{x} < y \ ) \,] \tag{14}$$

One difficulty with the $\text{IS}_D^\#(\beta)$ and $\text{IS}_D^\#(\beta_{A,i})$ systems was mentioned by Example 5.6. It was that although Peano Arithmetic can corroborate (8)'s invariance for every Π_1^* sentence Ψ, these latter two systems cannot always also do so.

While there will probably never be a perfect method for fully resolving this challenge, there is a pragmatic engineering-style solution that is often available. This is essentially because our proof of Theorem 5.7 employed a formalism β that used essentially only one axiom sentence (e.g. (12)'s Π_1^* declaration).

Since the $\text{IS}_D^\#(\beta)$ formalism was intended for use by any finite-sized system β, it is clearly possible to include any finite number of formally true Π_1^* sentences in β. Thus for some fixed constant L, one can easily let β include L copies of (12)'s axiom framework for a finite number of different Test_1, Test_2 ... Test_L predicates, each of which satisfy Definition 5.3's criteria for being kernelized formulae. In this case, $\text{IS}_D^\#(\beta)$ will formally map each initial Π_1^* theorem Ψ of some axiom system A onto L resulting different Π_1^* theorems of the form (11).

Our conjecture is that a goodly number of issues concerning some logic-based engineering applications, called say E, will have convenient solutions via self-justifying axiom systems that follow the strategy outlined in the preceding paragraph. Thus, we are suggesting that if β is a large-but-finite set of axioms, that consists of L copies of (12)'s axiom framework for different Test_1... Test_L predicates, then at least some futuristic engineering applications E may possibly have their needs met by an $\text{IS}_D^\#(\beta)$ formalisms, when a software engineer meticulously chooses [2] a proper finite-sized β that fits its needs.

---

[2] In addition to having L copies of (12)'s axiom framework, it will be probably preferable (but not necessary)

Remark 6.1 The preceding discussion was not meant to overlook the fact that the Second Incompleteness Theorem is a robust result, that will preclude self-justifying systems from recognizing their own consistency under strong definitions of consistency (as was discussed by Sections 2 and 3). Our suggestion, however, is that computers are becoming so powerful, in both speed and memory size as the 21st century is progressing, that there likely will emerge engineering-style applications E that will benefit from $\text{IS}_D^{\#}(\beta)$'s self-referencing formalisms when a *large-but-finite-sized* β is delicately chosen. (Moreover, it is of interest to speculate about whether such computers can, at least partially, imitate a human's gut instinctive *at least partial* confidence [3] in his own consistency.)

Remark 6.2 The next three sections will further amplify on the preceding themes. They will make it unequivocally clear that the Second Incompleteness Theorem is too powerful a result for any partial exception to overcome its formalism, in an entire and full sense. Yet at the same time, they will suggest that some aspects of the statements $*$ and $**$, by Gödel and Hilbert, did correctly foresee that a logical system can, at least, bolster a *partial limited-style appreciation* of its own consistency.

7   Comparing the Properties of Type-M and Type-A Formalisms

Let us recall that axioms (1)-(3) indicated that Type-A systems differ from Type-M formalisms by treating Multiplication as a 3-way relation (rather than as a total function). For the sake of accurately characterizing what our systems can and cannot do, we have described our results as being fringe-like exceptions to the Second Incompleteness Theorem from the perspective of an Utopian view of Mathematics, while perhaps being more significant results from an engineering-style perspective of knowledge. Our goal in this section will be to amplify upon this perspective by taking a closer look at Type-A and Type-M formalisms.

---

to have some finite number of say M additional general purpose Π_1^* axioms added to β, so as to enhance the power and increase the efficiency of $\text{IS}_D^{\#}(\beta)$.

[3] While human beings presumably do not have a fully robust conficence in their own consistency. the footnoted sentence refers to the fact that they certainly have a sufficient *partial* confidence of such to, at least, reach the threshold needed for motivating themselves to cogitate.

During our discussion, $x_0, x_1, x_2, ...$ and $y_0, y_1, y_2, ...$ will denote the sequences defined by the identifies of:

$$x_0 \;=\; 2 \;=\; y_0 \tag{15}$$

$$x_{i+1} \;=\; x_i \,+\, x_i \tag{16}$$

$$y_{i+1} \;=\; y_i \,*\, y_i \tag{17}$$

These sequences represent the growth rates that are produced, respectively, by the axioms (2) and (3), that specify addition and multiplication are total functions.

Equations (15)-(17) imply $y_n \;=\; 2^{2^n}$ and $x_n \;=\; 2^{n+1}$. Thus, the $y_0, y_1, y_2, ...$ sequence will grow at a much faster rate than the $x_0, x_1, x_2, ...$ sequence. (This is because y_n's binary encoding will have an $\text{Log}(y_n) \;=\; 2^n$ length while x_n's binary encoding will have a much smaller length of size $\text{Log}(x_n) \;=\; n + 1.$)

Our prior papers noted that the difference between these growth rates was the reason that [35, 37, 43] showed all natural Type-M systems, recognizing integer-multiplication as a total function, were unable to recognize their tableaux-styled consistency — while [34, 36, 40] showed some Type-A systems could simultaneously prove all Peano Arithmetic Π_1^* theorems and corroborate their own tableaux consistency. Their gist was that a Gödel-like diagonalization argument, which causes an axiom system to become inconsistent as soon as it proves a theorem affirming its own tableaux consistency, stems from the exponential growth in the series $y_0, y_1, y_2, ...$. (In other words, this growth facilitates an intense amount of self-referencing, using the identity $\text{Log}(y_n) \;\cong\; 2^n$, that will invoke the force of Gödel's seminal diagonalization machinery.)

These issues do raise the following question about proofs of the Second Incompleteness Theorems:

- How natural are exponentially growing sequences analogous to $y_0, y_1, y_2, ...$, whose $n-$th member requires more than 2^n bits to encode, when such encodings exceed the number of atoms in the universe if simply $n > 100$? Does the employment of such a sequence, for corroborating the Second Incompleteness Effect, suggest that this formalism is relying upon partially artificial constructs?

18

We will not attempt to derive a Yes-or-No answer to Question $\bullet$ because it is one of those epistemological questions that can be debated endlessly. Our point is that $\bullet$ probably does not require a definitive positive or negative answer because both perspectives are useful. Thus, the theoretical existence of a sequence integers of $y_0, y_1, y_2, ...,$ whose binary encodings are doubling in length, is tempting from the perspective of an Utopian view of mathematics, while awkward from an engineering styled perspective. We therefore ask: *"Why not be tolerant of both perspectives?"*

One virtue of this tolerance is it ushers in a greater understanding of the statements $*$ and $**$ that Gödel and Hilbert made during the first third of the 20th century. This is because the Incompleteness Theorem demonstrates no formalism can display an understanding of its own consistency in an idealized Utopian sense. On the other hand, §6 suggested these remarks by Gödel and Hilbert might receive more sympathetic interpretations, if one sought to explore such questions from a less ambitious almost engineering-style perspective.

Our main thesis is supported by a theorem from [41]. It indicated that semantic tableaux variations of self-justifying systems have no difficulty in recognizing that an infinitized generalization of a computer's floating point multiplication (with rounding) is a total function. The latter differs from integer-multiplication, by not having its output become double the length of its input when a number is multiplied by itself. Thus, the intuitive reason that [41]'s multiplication-with-rounding operation is compatible with self-justification is because it avoids the inexorable exponential growth, associated with Equation (17)'s sequence $y_0, y_1, y_2, ...$.

Also, the Theorem 7.1 below indicates our self-justifying system can recognize a double-precision variant of integer multiplication as a total function.

Theorem 7.1 *Consider the version of the $IS_D(A)$ and $IS_D^{\#}(\beta_{A,i})$ axiom systems where A denotes Peano Arithmetic. These systems can formalize two total functions, called Left(a, b) and Right(a, b), where any ordered pair (a, b) is mapped onto the objects, Left(a, b) and Right(a, b), designating the bit-sequences that represent the left and right halves of the multiplicative product*

19

of a and b.

Theorem 7.1 follows from a straightforward generalization of [41]'s analysis of floating point multiplication. Both it and the [41]'s result suggest that many arithmetic operations, appearing in an engineering environment, are nicely compatible with self justification. This is because most of the engineering and applied-mathematics uses of multiplication can replace a purist variant of integer multiplication with either floating point multiplication or Theorem 7.1's notion of double-precision multiplication.

8   A Different Type of Evidence Supporting Our Thesis

Let us recall that Pudlák and Solovay [26, 28] observed that essentially all Type-S systems, containing merely statement (1)'s axiom that successor is a total function, are unable to verify their own consistency under Hilbert deduction. (See also related work by Buss-Ignjatovic [6], Švejdar [30] and the Appendix A of [36])

It turns out that [39] generalized these results to show that Type-A systems are unable to verify their own consistency, under Definition 3.2's Tab-2 deduction methodology. At the same time, Theorems 4.1 and 5.7 demonstrate that the IS_D and $\text{IS}_D^{\#}$ axiomatic frameworks can verify their own consistency under Tab-1 deduction. Our goal in this section will be to illustrate how the contrast between these positive and negative results is analogous to the differing growth rates of the sequences $x_0, x_1, x_2, ...$ and $y_0, y_1, y_2, ...$ from Equations (15)–(17).

During our discussion $G_i(v)$ will denote the scalar-multiplication operation that maps an integer v onto $2^{2^i} \cdot v$. Also, Υ_i will denote the statement, in the U-Grounding language, that declares that G_i is a total function. Our paper [39] had noted that Υ_i can receive a Π_2^* encoding. It is also obvious that G_i satisfies the identity:

$$G_{i+1}(v) \quad = \quad G_i(\ G_i(v) \ ) \tag{18}$$

It was noted in [39] that this identity implies one can construct an axiom system β, comprised of solely Π_1^* sentences, where a semantic tableaux proof can establish Υ_{i+1} from $\beta + \Upsilon_i$ in a

constant number of steps. This implies, in turn, that a Tab-2 proof from β will require no more that $\mathrm{O}(n)$ steps to prove Υ_n (when it uses the obvious n-step process to confirm in chronological order Υ_1, Υ_2, ...Υ_n .)

These observations are significant because $G_n(1) = 2^{2^n}$. Thus, [39] showed a Tab-2 proof from β can verify in $\mathrm{O}(n)$ steps that this integer exists.

This example is helpful because it illustrates that the difference between the growth rates, under proofs using Definition 3.2's Tab-1 and Tab-2 deductive methodologies, is analogous to the differing growth speeds of the sequences $x_0, x_1, x_2, ...$ and $y_0, y_1, y_2, ...$ from Equations (15)–(17). Thus once again, the faster growth-rate has the side-effect of triggering off the force of the Second Incompleteness Theorem (e.g. see [39]).

This analogy suggests that the Second Incompleteness Theorem has different implications from the perspectives of Utopian and engineering theories about the intended applications of mathematics. Thus, a Utopian may possibly be comfortable with a perspective, that contemplates sequences $y_0, y_1, y_2, ...$ with elements growing in length at an exponential speed, but many engineers may be suspicious of such growths.

A hard-core engineer, in contrast, might surmise that the inability of self-justifying formalisms to be compatible with Definition 3.2's Tab-2 deduction is not as disturbing as it might initially appear to be. This is because Tab-2 differs from Tab-1 deduction by producing growth rates that are so sharp that their material realization has no analog in the everyday mechanical reality that is the focus of an engineer's interest.

Our personal preference is for a perspective lying half-way between that of an Utopian mathematician and a hard-nosed engineer. Its dualistic approach suggests some form of half-way agreement with the the goals of Hilbert's consistency program in $**$. (This is the maximal type of agreement with Hilbert's goals that is, obviously, feasible because no more than a curtailed form of the objectives of his consistency program is plainly realistic.)

9   Further Related Improvements

An added point is that there are many types of self-justifying systems available, with some being better suited for engineering environments than others.

For instance, our initial 1993 paper [34] employed a Group-3 *"I am consistent"* axiom that was much weaker than the current specimen. The distinction was that [34]'s self-consistency declaration excluded merely the existence of a semantic tableaux proof of $0 = 1$ from itself, while the sentence (7) is more elaborate because it excludes the existence of simultaneous proofs of a Π_1^* theorem and its negation.

This distinction is significant because it implies that each time our self-justifying systems S prove a Π_1^* or Σ_1^* theorem Φ , they will know it is fruitless to search for a proof of the negation of Φ's statement [4].

Ideally, one would like to additionally develop self-justifying systems S that could corroborate the validity of (19)'s reflection principle for all sentences Φ, as well.

$$\forall p \; [\; Prf_S^D(\ulcorner \Phi \urcorner, p) \;\; \Rightarrow \;\; \Phi \;] \tag{19}$$

Löb's Theorem establishes, however, that all axiom systems S, containing the strength of Peano Arithmetic, are able to prove (19)'s invariant *only in the degenerate case* where they prove Φ itself. Also, the Theorem 7.2 from [36] showed essentially all axiom systems, *weaker* than Peano Arithmetic, are unable to prove (19) for all Π_1^* sentences Φ simultaneously. Theorem 9.1's reflection principle is thus pleasing:

Theorem 9.1 *For any input axiom system A, it is possible to extend the self-justifying $IS_D(A)$ and $IS_D^{\#}(\beta_{A,i})$ systems, from Theorems 4.1 and 5.7, so that a new broader system S has all the properties of the prior systems and can additionally:*

---

[4] From at least an engineer's hard-nosed perspective, this fact is very helpful because it will allow an automated theorem prover to imitate, for at least Π_1^* and Σ_1^* sentences, the common-sense human presumption that the discovery of a proof a theorem Φ leads to the time-saving conclusion that it is unnecessary to search for a proof of $\neg \Phi$.

1. *Verify that* Tab-1 *deduction supports the following analog of (19)'s self-reflection principle under S for any Δ_0^* and Σ_1^* sentences* Φ :

$$\forall p \ [\ Prf_S^{\text{Tab}-1}(\ulcorner\Phi\urcorner, p) \ \Rightarrow \ \Phi \ ] \tag{20}$$

2. *Verify (21)'s more general* **"root-diluted"** *reflection principle for* S *whenever θ is Σ_1^* and Φ is a Π_2^* sentence of the form* " $\forall u_1...\forall u_n \ \ \theta(u_1...u_n)$ ".

$$\forall p \ [\ Prf_S^{\text{Tab}-1}(\ulcorner\Phi\urcorner, p) \ \Rightarrow \ \forall x \ \ \forall u_1 < \sqrt{x} \ \ ... \ \ \forall u_n < \sqrt{x} \ \ \theta(u_1...u_n) \ ] \tag{21}$$

Theorem 9.1 is stronger than Theorems 4.1 and 5.7 because of its reflection principles. Its proof is a generalization of related reflection princibles used in [36]. It is summarized in Appendix B. Theorem 9.1 reinforces our theme about how exceptions to the Second Incompleteness Theorem may appear to be minor from the perspective of an Utopian view of mathematics, while being significant from an engineering standpoint. This is because:

A. The ability of Theorem 9.1's system S to support (20)'s self-reflection principle under Tab-1 proofs for arbitrary Δ_0^* and Σ_1^* sentences, as well as to support (21)'s root-diluted reflection principle for Π_2^* sentences, is clearly non-trivial.

B. The incompleteness result in [36]'s Theorem 7.2, shows, however, that the above reflection principle cannot be extended to all Π_1^* sentences, *in an undiluted sense.*

The tight fit between A and B is analogous to several other slender borderlines that separated generalizations and boundary-case exceptions for the Second Incompleteness Theorem, that we had mentioned earlier. The Second Incompleteness Theorem should, thus, be seen as robust, from an idealized Utopian perspective on mathematics, while permitting caveats from possibly an engineering styled perspective. This dualistic perspective allows one to nicely share at least some *partial agreement* with the general spirit of the the statements $*$ and $**$ by Gödel and Hilbert, while simultaneously appreciating the stunning achievement of the Second Incompleteness Theorem.

10 "Wir müssen wissen: wir werden wissen"

In summary, our research has been largely motivated by the approximate conjecture that human beings acquire the needed mental energy and will-power for motivating their time-consuming cogitations *only via* the human mind owning, *at least in a weak sense*, some type of quasi-automatic and possibly unconscious appreciation of its own consistency.

Two of Hilbert's famous often-quoted statements suggest that he would agree, at least partially, with the main spirit of our conjectures. The first was his 1925 statement $**$ (whose relationship to Gödel's supportive remarks in $*$ was already noted in §1). The second Hilbert statement was the widely known motto of his consistency program, "Wir müssen wissen: wir werden wissen", whose English translation appears below. Part of what makes $***$ intriguing is that Hilbert arranged for this motto, often cited as the justification of his consistency program, to be placed on his tombstone, even though he co-authored with Bernays an historic generalization of the Second Incompleteness Theorem, using their ubiquitous Hilbert-Bernays Derivability Conditions [17, 23].

$***$ "We must know: we will know"

The presumable reason that Hilbert believed that, at least, some diluted version of his consistency program would ultimately succeed is because it adamantly defies common sense to explain how a human mind can motivate itself to cogitate, without *possessing some at least diluted* form of knowledge and faith in its own consistency.

Thus, Hilbert clinged to the belief that some *at least diluted version* of his consistency program would ultimately succeed, as $***$ had hinted [5]. Moreover, a word-by-word reading of

---

[5] Some readers may initially suspect that the "Wir müssen wissen: wir werden wissen" statement was inscribed on Hilbert's tombstone to make a general statement about his philosophy of life, rather than to address remaining issues left open by Gödel's Second Incompleteness Theorem. While this point may be partially true, the significance of $***$ being inscribed on his tombstone seems to transcend this viewpoint. This is because Hilbert was famous for having adamantly continued, throughout his life, seeking a "revised" version of his consistency program, even when he co-authored [17]'s generalization of the Second Incompleteness Theorem. Thus, it appears

∗ ∗ ∗ finds it openly declaring some mysterious object XYZ "must" exist, whose exact nature (and even defining name) it is deliberately vague about (despite the fact that Hilbert tells us that "we must know" about it).

The prior nine sections of the current article, as well as our earlier research in [34]-[45], were largely devoted to defining this elusive object XYZ (that underlines a human's instinctive faith in his own processes but which is decisively awkward to describe). Part of the reason that Hilbert's desired XYZ is so tenderly elusive is that §7 has documented that systems, recognizing *even* their own *tiny* miniaturized tableaux consistency, cannot corroborate the assumption that multiplication is a total function, and §8 has formalized how Type-A axiom systems are also unable to recognize their own consistency under a Tab−2 generalization of semantic tableaux deduction.

These facts certainly illustrate that a formalism cannot recognize its own consistency in an idealized Utopian sense, and one must thus approach this subject matter with daring and ginger amounts of caution.

Within this context, the gist of our Sections 7 and 8 was that Type-A systems that recognize their own consistency under Tab−1 deduction are not, actually, quite as weak as they might initially appear. This is because they avoid the super-exponential growth of Equation (17)'s $y_0, y_1, y_2, ...$ series, whose growth is inscrutable from a hard-nosed engineering style perspective. Thus, if one confines one's attention onto Equation (16)'s more modest exponential growth for its sequence $x_0, x_1, x_2, ...$, then self-justifying systems can be formalized that are compatible with this slower growth rate and meet at least a diluted version of ∗ ∗ ∗'s objectives. (These weak-styled systems, obviously, support only a *tenderly modest* knowledge of their own self-consistency, relative to an idealistic Utopian's aspirations for what mathematics should ideally embody. Their outlook about self-consistency are, however, *not entirely* inconsequential from an engineer's applications-oriented perspective.)

---

to be no coincidence that Hilbert, who had a knack for formulating many famous often-quoted phrases, would have the motto of his never-terminated consistency program inscribed as the final memorial on his tombstone.

The thesis, advanced in this article, is subtle because there are many different kinds of *"I am consistent"* statements that can be used by the Group-3 components of our self-justifying axiom systems. Some of these formalisms have the advantages of being preferable from a pedagogic perspective. This is because a formalism that merely declares the non-existence of a semantic tableaux proof of $0 = 1$ from itself is relatively easy to analyze (such as the initial 1993 version of a self-justifying axiom system that was proposed in [34]). Other variants of the self-justification construct, which assert the non-existence of simultaneous Tab-1 styled proofs of a Π_1^* sentence and its negation, are substantially more complex to analyze but have the advantage of providing greater levels of self-knowledge. This second topic, obviously, comes closer to meeting Hilbert's hopes, subsequent to 1931, because its self-justifying formalisms contain a larger degree of inner strength and self-understanding.

Finally, it is useful to conclude our perspective about a limited partial revival of Hilbert's consistency program by reviewing the contrast between Theorems 4.1 and 5.7. These theorems formalize a fundamental trade-off, where each result has sharply different advantages. Thus, Theorem 4.1 indicates that every consistent axiom system A can be mapped onto a recursively enumerable self-justifying axiom system $\text{IS}_D(A)$ that can formalize all A's Π_1^* knowledge while recognizing its own Level-1 consistency. In contrast, Theorem 5.7 shows that the infinite number of different proper axioms in $\text{IS}_D(A)$ can be reduced to a finite size if one is willing to settle for a system $\text{IS}_D^{\#}(\beta_{A,i})$ that replaces a full knowledge of all A's Π_1^* theorems with an $i-$th kernelized knowledge of this information. Since part of Hilbert's goal was to find a means whereby systems *of purely finite size* can recognize their own consistency, we suspect that Theorem 5.7, along with Section 6's enhancement of its results, demonstrate that some fragmentary (but not full) part of Hilbert's objectives in $**$ can be achieved.

11 Miniaturized Finitism

The phrase *"Miniaturized Finitism"* should, perhaps, be employed to describe the type of agenda for mathematics that this article is advocating. The cautious-sounding adjective of *"miniaturized"* was attached to this name because one should readily admit that there is an indisputably miniaturized and quite humble aspect to any school of mathematics that drops the assumption that multiplication is a total function and which also replaces an unqualified modus ponens rule with Definition 3.2's more tender Tab−1 variant of deduction.

The second word *"Finitism"*, on the other hand, was attached to this name because it partially vindicates the "finistic" presumption of Gödel and Hilbert, specified in statements ∗ and ∗∗, that it should be feasible to formalize how Thinking Beings do conceptualize at least a partial *intuitive* appreciation of their own consistency. Thus, Miniaturized Finitism looks towards the remaining approximate 10 % of the issues raised by Hilbert's Second Open Question, in a context where the Second Incompleteness Theorem has resolved the more main-stream 90 % of the issues, raised by Hilbert's year-1900 open question, in a decisively negative manner [6].

Some may critically view our particular form of self-justification as an essentially dwarfed-sized theory, which uses an argument resting on a tiny "miniaturized" notion of growth to explain how Thinking Beings are able to muster an at least partial appreciation of their own consistency. In such a context, it is pleasing to reply that Gödel was famous for telling Einstein that he did dearly *love* dwarfs: For instance, Dawson [7] and Goldstein [14] recall how Gödel "was espeically fond of Disney films" and attended "at least three" repeated screenings of "Snow White and the Seven Dwarfs". Likewise, Yourgrau [46] recites Gödel telling Einstein that the

---

[6] Researchers who happen to examine Willard's research since 1990 will curiously notice that the philosophy of the footnoted sentence has been central to this research since 1990. Thus, Fredman-Willard [10, 11] published in that year a boundary-case exception to the widely accepted textbook lower bound for sorting and searching that would later become the chronologically first among six items listed in the *Mathematics and Computer Science* section of the *National Science Foundation's 1991 Annual Report.* And so beginning with [34]'s initial 1993 theorem about self verification, Willard's on-going research into Logic was based on the presumption that other types of widely-accepted enticing textbook results would be found, upon careful inspection, ultimately to admit partial, *although certainly weak,* forms of *very tightly* defined boundary-case exceptions.

tale of the seven dwarfs was, actually, his favorite movie because it "presents the world as it should be and as if it had meaning".

Leaving aside the poetry of the preceding metaphor about how Gödel was enamored by miniaturized-sized objects such as dwarfs, our main points are that both Gödel and Hilbert specified in statements $*$ and $**$ that they felt some type of self-justification would be feasible. Thus, the name *"Miniaturized Finitism"*, has its adjective *"miniaturized"* nicely capturing the fact that only a discernible *fractional* (but indeed quite non-trivial) part of the initial goals of Hilbert's consistency program are achievable.

12 Summary Statement

As a final note, it is useful to recall that the opening chapter of this article indicated the current report would have different goals than our earlier papers.

The distinction was that our earlier work consisted of papers employing mostly a mathematical gendre, focusing on generalizations and boundary-case exceptions for the Second Incompleteness Theorem. The emphasis in the current report was different because it sought to focus on interpreting the meaning of the statements $*$, $**$ and $***$ of Gödel and Hilbert, rather than centering around the mathematics of our new formalisms. (Our current paper presented new theorems only when they were necessary to formalize its epistemological overview.)

The reason for the current article's different emphasis is that we anticipate that this second topic should grow in significance, as the 21st century progresses. This is mostly because digital computers have grown in importance during the last seventy years and have essentially doubled in memory size every 2 years at the unabated exponential rate that Moore's Law Moore's Law has predicted.

This fact about the burgeoning growth in memory size for digital computers in turn raises the questions about their other capacities. In particular, it raises the question of whether such

computers will have the ability to in some sense to recognize their own consistency, which ability humans seem to possess in some partial, though not full sense.

Regarding the last point, our speculations are not intended to suggest that humans (or computers) will be able to corroborate their own consistency in a robust sense. It is obvious that the Second Incompleteness Theorem has established that such a prospect is hopeless. The point is, however, that humans have been able to intuitively appreciate their own consistency in some type of quasi-reasonable sense — in order to motivate themselves to gain the mental energy and will power to cogitate.

The advance of the digital computer during the 21st century will, thus, inevitably raise the question about how computers can maintain, likewise, some type of timid-but-well-defined appreciation of their own consistency. We suspect this will cause the statements $*$, $**$ and $***$ of Gödel and Hilbert to be reviewed often, again, in the future (and ultimately to gain some form of partial-although-not-full reinforcement).

This point is delicately subtle because there is no question that the Second Incompleteness Theorem will always be remembered as the greatest of all the achievements in the modern advancement of Logic. Yet, this fact should not preclude the Second Incompleteness Theorem as being seen as analogous to many other mathematical discoveries, in that most great mathematical results typically permit for the existence of some types of well-defined forms of boundary-case exceptions.

Thus along the lines that Gödel and Hilbert had approximately predicted, in especially their first and third statements of $*$ and $***$, it is not surprising that there exists certain types of unconventionally framed axiom systems that possess a form of diluted *but not fully immaterial* knowledge about their own self-consistency.

29

Appendix A summarizing Theorem 5.2's Proof

The same methods that Section 5 of [40] had used to prove Theorem 4.1 can be easily generalized to corroborate Theorem 5.2's similar consistency preservation result. This is because Theorem 5.2's hypothesis implies that all the Group 0, 1 and 2 axioms of $\mathrm{IS}_D^{\#}(\beta)$ are analogous to their counterparts under $\mathrm{IS}_D(A)$, in that both represent Π_1^* sentences holding true under the standard model of the Natural Numbers. In this context, a similar reductionist argument, as had appeared in Section 5 of [40], will imply $\mathrm{IS}_D^{\#}(\beta)$'s Group-3 axiom must also hold true. Thus, a direct analog of [40]'s syllogism will show that it is impossible for an ordered pair (p, q) to contradict $\mathrm{IS}_D^{\#}(\beta)$'s Group-3 axiom by simultaneously having 1) p prove a Π^* theorem, 2) q prove its negation and 3) Max(p, q) represent the minimal value among all ordered pairs with this property.

We will not delve into further details here because Theorem 5.2's proof is similar to [40]'s analogous theorem in that they both, essentially, rest on the fact that the Group 0, 1 and 2 axioms of their formalism represent true Π_1^* sentences. The reason for our interest in Theorem 5.2 in the current article is that it is a useful intermediate step for establishing Theorem 5.7. The latter, in turn, will reinforce the epistemological viewpoint of Sections 6 – 12 (especially as it pertains to the question about whether a limited subset of Hilbert's "finistic goals" can be revived in a fragmentized sense).

Appendix B summarizing Theorem 9.1's Proof :

Let Φ denote a prenex sentence of the form $\forall u_1\ \exists w_1\ \forall u_2\ \exists w_2...\forall u_n\ \exists w_n\ \ \theta(u_1, w_1...u_n, w_n)$ where $\theta(u_1, w_1...u_n, w_n)$ is a Δ_0^* formula. Also, let us define $\Phi^{\oplus}$ to be the following sentence:

$$\forall x\quad \forall u_1 < \sqrt{x}\quad \exists w_1\quad \forall u_2\ < \sqrt{x}\quad \exists w_2\ .....\ \forall u_n < \sqrt{x}\quad \exists w_n\quad \theta(u_1, w_1...u_n, w_n) \tag{22}$$

This notation was designed so that the operation that maps Φ onto $\Phi^{\oplus}$ will change the ranges *of only* the unbounded universal quantifiers in Φ (e.g. all *bounded* universal quantifiers appearing inside its expression θ are left unchanged). Thus, $\Phi\ =\ \Phi^{\oplus}$ in the degenerate case where Φ is either a Δ_0^* or Σ_1^* sentence.

An axiom system S will be said to own a "tangible understanding" of its reflection properties under deduction method D iff it can corroborate (23) for every sentence Φ.

$$\forall p \ [\ Prf_S^D(\ulcorner \Phi \urcorner, p) \ \Rightarrow \ \Phi^\oplus \ ] \tag{23}$$

The Theorem 6.1 of [36] showed that every consistent axiom system A can be mapped onto a system S that can prove all A's Π_1^* theorems, while verifying (23)'s reflection principle when D denotes semantic tableaux deduction.

Moreover, it is possible to hybridize [36]'s result with Theorem 5.7's methodology to establish that more advanced versions of S can verify (23) for the cases where 1) D denotes the more elaborate $Tab-1$ *(rather than semantic tableaux)* deduction method, 2) S can possess, for any fixed i, a finite-sized Group-2 scheme that enables it to prove the $i-$th kernelized image of each Π_1^* theorem of A, and 3) S can verify its own Level-1 consistency.

Such an argument will corroborate Theorem 9.1's claims about (20) and (21)'s reflection principles. (For instance, it applies to (20) because $\Phi \ = \ \Phi^\oplus$ for all Δ_0^* or Σ_1^* sentences.)

We will not provide more details concerning Theorem 9.1's proof here because the overall structure of its justification is analogous to [36]'s corroboration of its "TangRoot" reflection principle. It is desirable to keep this technical report's mathematical discussion as brief as possible, so that our discourse can focus, instead, on an epistemological interpretation of the significance of self justification.

References

[1] Z. Adamowicz, "Herbrand Consistency and Bounded Arithmetic", *Fundamenta Mathematica* 171 (2002) pp. 279-292.

[2] Z. Adamowicz and P. Zbierski, "On Herbrand consistency in weak theories", *Archive for Mathematical Logic* 40 (2001) pp. 399-413.

[3] A. Bezboruah and J. Shepherdson, "Gödel's Second Incompleteness Theorem for Q", *Journal of Symb Logic* 41 (1976) 503-512.

[4] S. Bringsjord, private communications during 1994 suggesting that I publish my results about self-justifying logics in two stages, with its first phase presenting a series of mathematical results and second stage providing a philosophical interpretation of such results. Selmer thought this course was preferable because philosophical interpretations always lend themselves to wider levels of debates. Thus following Bringsjord's suggestion, we had postponed discussing the philosophical interpertation of [34]-[45]'s mathematical results until the current report.

[5] S. R. Buss, *Bounded Arithmetic,* (Ph D Thesis) Proof Theory Notes #3, Bibliopolis 1986.

[6] S. R. Buss and A. Ignjatovic, "Unprovability of Consistency Statements in Fragments of Bounded Arithmetic", *Annals of Pure and Applied Logic* 74 (1995) pp. 221-244.

[7] John W. Dawson Jr. *Logical Dilemmas: The Life and Work of Kurt Gödel,* A.K. Peter Press, 1997.

[8] H. B. Enderton, *A Mathematical Introduction to Logic*, Acadmic Press, 2001.

[9] M. Fitting, *First Order Logic and Automated Theorem Proving,* SpringerVerlag 1996.

[10] M. L. Fredman and D. Willard, Blasting Through the Information Theoretic Barrier with Fusion Trees, in *Proceedings of 22nd ACM Symposium on the Theory of Computing (1990)* pp. 1–7 (whose abridged proof-sketches were later published by [11] in a more detailed form).

[11] M. L. Fredman and D. Willard, Surpassing the Information Theoretic Barrier with Fusion Trees, *Journal of Computer and Systems Sciences,* 47 (1993) pp. 424–436.

[12] H. M. Friedman, "Translatability and Relative Consistency", Ohio State Tech Report, 1979.

[13] K. Gödel, " Über formal unentscheidbare Sätse der Principia Mathematica und Verwandte Systeme I", *Monatshefte für Math. Phys.* 37 (1931) pp. 349-360.

[14] R. Goldstein, *Incompleteness The Proof and Paradox of Kurt Gödel*, Atlas Books of W. Norton and Company, 2005.

[15] P. Hájek and P. Pudlák, *Metamathematics of First Order Arithmetic,* Springer Verlag 1991.

[16] D. Hilbert, Uber das Unendliche, Springer 1925. (English Translation (1967) by J V Heijenoort in *From Frege to Gödel: A Source Book in Math Logic,* Harvard Press, page 375)

[17] D. Hilbert and P. Bernays, *Grundlagen der Mathematic*, Springer 1939.

[18] R. Jeroslow, "Consistency Statements in Formal Mathematics", *Fund Math* 72(1971) 17-40.

[19] S. C. Kleene, "On the Notation of Ordinal Numbers", *Journal Symb Logic* 3 (1938), 150-156.

[20] L. A. Kołodziejczyk, Private Email Communications during November of 2005.

[21] G. Kreisel and G. Takeuti, "Formally Self-Referential Propositions for Cut-Free Classical Analysis", *Dissertationes Mathematicae* 118 (1974) pp. 1–55

[22] M. H. Löb, A Solution to a Problem by Leon Henkin, *Jour Symb Logic* 20(1955) 115-118

[23] E. Mendelson, *Introduction to Mathematical Logic*, CRC Press, 2010.

[24] E. Nelson, *Predicative Arithmetic,* Princeton Math Notes, 1986.

[25] R. Parikh, "Existence and Feasibility in Arithmetic", *Journal Symb Logic* 36 (1971) 494-508

[26] P. Pudlák, "Cuts, Consistency Statements&Interpretations", *Journal Symb Logic* 50(1985)423-442

[27] H. Rogers Jr. , *Recursive Functions and Effective Compatibility,* McGrawHill 1967.

[28] R. M. Solovay, Private telephone conversations in 1994 describing Solovay's generalization of one of Pudlák's theorems [26], using some methods of Nelson and Wilkie-Paris [24, 33]. (The Appendix A of [36] is a 4-page approximate reconstruction of this conversation.)

[29] Stanford Encyclopedia Entry on "Kurt G"odel, See the second pargraph of its Section 2.2.4 for a description of Gödel's reaction to Turing's paper [31].

[30] V. Švejdar, "An Interpretation of Robinson Arithmetic in its Grzegorczjk's Weaker Variant" *Fundamenta Informaticae* 81 (2007) pp. 347-354.

[31] A. M. Turing, "On Computable Numbers with an Application to the Entscheidungsproblem", Proceedings of the London Mathematical Soceity (Series 2), 42 (1937) pp. 230-265.

[32] A. Visser, "Faith and Falsity", *Annals of Pure and Applied Logic* 131 (2005) pp. 103-131.

[33] A. J. Wilkie and J. B. Paris, "On the Scheme of Induction for Bounded Arithmetic", *Annals of Pure and Applied Logic* (35) 1987, 261-302

[34] D. Willard, "Self-Verifying Axiom Systems", *Proceedings of the Third Kurt Gödel Colloquium* (1993), Springer-Verlag LNCS#713, pp. 325-336.

[35] —, "The Semantic Tableaux Version of the Second Incompleteness Theorem Extends to Robinson's Arithmetic Q", *Proceedings of Tableaux 2000 Conf*, SpringerVerlag LNAI#1847, 415-430

[36] —, "Self-Verifying Systems, the Incompleteness Theorem and the Tangibility Principle", in *Journal of Symbolic Logic* 66 (2001) pp. 536-596.

[37] —, "How to Extend The Semantic Tableaux And Cut-Free Versions of the Second Incompleteness Theorem Almost to Robinson's Arithmetic Q", in *Journal of Symbolic Logic* 67 (2002) pp. 465–496.

[38] —, "Some Exceptions for the Semantic Tableaux Version of the Second Incompleteness Theorem", *Proceedings of Tableaux 2002 Conf*, SpringerVerlag LNAI#2381, pp. 281–297.

[39] —, "A Version of the Second Incompleteness Theorem For Axiom Systems that Recognize Addition But Not Multiplication as a Total Function", *First Order Logic Revisited, (Year 2003 Proceedings FOL-75 Conference),* Logos Verlag (Berlin) 2004, pp. 337–368.

[40] —, "An Exploration of the Partial Respects in which an Axiom System Recognizing Solely Addition as a Total Function Can Verify Its Own Consistency", *Journal of Symbolic Logic* 70 (2005) pp. 1171-1209.

[41] —, "On the Available Partial Respects in which an Axiomatization for Real Valued Arithmetic Can Recognize its Consistency", *Journal of Symbolic Logic* 71 (2006) pp. 1189-1199.

[42] —, "A Generalization of the Second Incompleteness Theorem and Some Exceptions to It". *Annals of Pure and Applied Logic* 141 (2006) pp. 472-496.

[43] —, "Passive induction and a solution to a Paris-Wilkie question", *Annals of Pure and Applied Logic* 146(2007) pp. 124-149.

[44] —, "Some Special Axiomizations for IΣ_0 Manage to Evade the Herbrandized Version of the Second Incompleteness Theorem", *Information and Computation* 207(2009) 1078-1093

[45] — , "A Detailed Examination of Methods for Unifying, Simplifying and Extending Several Results About Self-Justifying Logics", http://arxiv.org/abs/1108.6330.

[46] P. Yourgrau, *A World Without Time: The Forgotten Legacy of Gödel and Einstein*, Basic Books (2005). (Our particular quotations from [46], appearing in §1, can be found on its page 58.)