A consistent formal system which verifies its own consistency

Nik Weaver

Washington University in St. Louis

9/11/2015

Truth and Assertibility, World Scientific Press, 2015

Assertibility

a sentence is *assertible* (*informally provable*, *constructively true*) if it can be affirmed on logically conclusive grounds

example: there are infinitely many prime numbers

Assertibility versus formal proof

does assertibility coincides with formal provability in PA (or ZFC)?

if we know that every theorem of PA is genuinely assertible, then we can infer that $0{=}1$ is not a theorem of PA, and thus we can informally — but conclusively — deduce an arithmetical sentence, $\mathrm{Con}(\mathrm{PA})$, which is not provable in PA

this is a general phenomenon: any formal system which is proposed to model arithmetical assertibility is either too strong (not known to be sound) or too weak (known to be sound, and therefore consistent)

we have an open-ended ability to go beyond any given formal system

Intuitionistic logic

the open-ended nature of assertibility (that we can go beyond any given formal system) suggests intuitionistic logic

- classical slogan: facts exist independently of whether they can be known

  $$A \vee \neg A$$

  (A or not-A: *law of excluded middle*)

- constructive slogan: there is no such thing as an unknowable fact

  $$A \to \mathbb{A}(A)$$

  (if A, then A is assertible: *capture law*)

The release law

$\mathbb{A}(A) \to A$ (*release law*) seems obvious: if we have definitive grounds to affirm A, then A must be true

this is fine if we have a classical truth predicate, but intuitionistic justification is circular:

a proof of "$\mathbb{A}(A) \to A$" manifests as a construction which converts any proof of the premise into a proof of the conclusion

i.e., it turns any proof that there is a proof of A into a proof of A

i.e., when given a construction of an object p and a proof that p proves A, it returns a proof of A — presumably, the object p

but this only works if we can infer that p proves A from a proof that p proves A; that is, we need to know $\mathbb{A}(p \vdash A) \to (p \vdash A)$

The assertible liar paradox

consider the sentence $L =$ "the negation of this sentence is assertible"

so $L \leftrightarrow \mathbb{A}(\neg L)$, and we can argue as follows

assume L
then $\mathbb{A}(\neg L)$
so $\neg L$ by release

this proves $\neg L$; that is, we have shown that $\neg L$ is assertible

but then we have proven L, contradiction on its face, *this appears to be a logically definitive argument which establishes* $L \wedge \neg L$

so if we take assertibility seriously, we should take the possibility that falsehoods are provable seriously

thus we should not assume the release law $\mathbb{A}(A) \rightarrow A$

but then the assertible liar paradox breaks down

A logical tightrope

if we assume falsehoods are not provable, then we can prove a falsehood

however, the paradoxical argument cannot be made if we do not assume this

the way to avoid proving falsehoods is to leave open the possibility that falsehoods may be provable

Axiomatizing assertibility

(treating $\mathbb{A}$ as a logical operator)

axiom scheme (C): $A \rightarrow \mathbb{A}(A)$

axiom ($\wedge$): $\mathbb{A}(A) \wedge \mathbb{A}(B) \rightarrow \mathbb{A}(A \wedge B)$

axiom ($\rightarrow$): $\mathbb{A}(A) \wedge \mathbb{A}(A \rightarrow B) \rightarrow \mathbb{A}(B)$

(alternatively, we can take $\mathbb{A}$ to be a unary predicate symbol which applies to Gödel numbers of sentences)

Some easy results

Prop. *If* $A_1 \wedge \cdots \wedge A_n \rightarrow B$ *then*

$$\mathbb{A}(A_1) \wedge \cdots \wedge \mathbb{A}(A_n) \rightarrow \mathbb{A}(B).$$

Corollary. *Let* A *and* B *be sentences and let* $C[x]$ *be a formula with one free variable* x. *Then*

- $\mathbb{A}(A \wedge B) \leftrightarrow \mathbb{A}(A) \wedge \mathbb{A}(B)$
- $\mathbb{A}(A) \vee \mathbb{A}(B) \rightarrow \mathbb{A}(A \vee B)$
- $\mathbb{A}(A \rightarrow B) \rightarrow (\mathbb{A}(A) \rightarrow \mathbb{A}(B))$
- $\mathbb{A}((\forall x)C[x]) \rightarrow \mathbb{A}(C[t])$
- $\mathbb{A}(C[t]) \rightarrow \mathbb{A}((\exists x)C[x])$

for any constant term t.

More on the assertible liar

recall $L \leftrightarrow \mathbb{A}(\neg L)$

$L \rightarrow \mathbb{A}(L)$ by capture, and $L \rightarrow \mathbb{A}(\neg L)$ by definition
so $L \rightarrow \mathbb{A}(L) \wedge \mathbb{A}(\neg L)$
so $L \rightarrow \mathbb{A}(\bot)$

now assume $\neg L$
then $\mathbb{A}(\neg L)$ by capture, and therefore L, contradiction
conclude $\neg\neg L$

we can prove $L \rightarrow \mathbb{A}(\bot)$ and $\neg\neg L$

the liar sentence entails that a falsehood is assertible, and its negation is false what if we assume the release law?

then from $L \to \mathbb{A}(\bot)$ infer $L \to \bot$, i.e., $\neg L$; together with $\neg\neg L$, this yields a contradiction

what if we assume LEM?

then we have $L \vee \neg L$, but we know $L \to \mathbb{A}(\bot)$ and $\neg\neg L$, i.e., $\neg L \to \bot$; so this yields $\mathbb{A}(\bot)$

$\mathrm{PA}^{\mathbb{A}}$: arithmetic with an assertibility predicate

intuitionistic logic plus LEM for all arithmetical formulas; peano axioms including induction for all formulas of the language

assertibility axioms:
(C): $A \to \mathbb{A}[\langle A \rangle]$ for every sentence A
($\wedge$): $\mathbb{A}[\langle \overline{A} \rangle] \wedge \mathbb{A}[\langle \overline{B} \rangle] \to \mathbb{A}[\langle \overline{A \wedge B} \rangle]$
($\to$): $\mathbb{A}[\langle \overline{A} \rangle] \wedge \mathbb{A}[\langle \overline{A \to B} \rangle] \to \mathbb{A}[\langle \overline{B} \rangle]$
($\forall$): $(\forall n)\mathbb{A}[\langle \overline{A[\hat{n}]} \rangle] \to \mathbb{A}[\langle \overline{(\forall n)A[n]} \rangle]$
(I): $\mathrm{Ax}[\langle A \rangle] \to \mathbb{A}[\langle A \rangle]$

$\overline{A}$ is the universal closure of a formula A and $\langle B \rangle$ is the Gödel number of a sentence B
$\mathrm{Ax}[n]$ is a primitive recursive formula which holds if n is the Gödel number of the universal closure of any axiom besides (I)

Some results

Theorem. *None of* $0 = 1$, $\mathbb{A}[\langle 0 = 1\rangle]$, $\mathbb{A}[\langle \mathbb{A}[\langle 0 = 1\rangle]\rangle]$, *etc., are theorems of* $\mathrm{PA}^{\mathbb{A}}$ *(complete consistency).*

Theorem. $\mathrm{PA}^{\mathbb{A}}$ *proves*

$$(\forall n)(\mathrm{Prov}_{\mathrm{PA}^{\mathbb{A}}}[n] \rightarrow \mathbb{A}[n])$$

(assertible soundness).

Theorem. $\mathrm{PA}^{\mathbb{A}}$ *proves*

$$\mathbb{A}[\langle \mathrm{Con}(\mathrm{PA}^{\mathbb{A}})\rangle]$$

(assertible consistency).

complete consistency is proven by constructing an upside-down Kripke model

assertible soundness is proven by a straightforward induction on the length of a proof

for assertible consistency, first prove in $\mathrm{PA}^{\mathbb{A}}$ that

$$(\forall n)\mathrm{Prov}_{\mathrm{PA}^{\mathbb{A}}}[\langle\neg\mathrm{Proof}[\hat{n},\langle 0=1\rangle]\rangle]$$

(for all n, $\mathrm{PA}^{\mathbb{A}}$ proves that n is not the Gödel number of a proof in $\mathrm{PA}^{\mathbb{A}}$ of 0=1), then use assertible soundness and the $(\forall)$ axiom to infer

$$\mathbb{A}[\langle(\forall n)\neg\mathrm{Proof}[\hat{n},\langle 0=1\rangle]\rangle]$$