About Truth and Types

Andrea Cantini

Dedicated to Gerhard Jäger on occasion of his 60th birthday.

Abstract We investigate a weakening of the classical theory of Frege structures and extensions thereof which naturally interpret (predicative) theories of explicit types and names à la Jäger.

1 Introduction

Non-extensionality is a basic feature of the so-called systems of *explicit mathematics* (EM in short), be they formalized in the style of Feferman [9] or in the framework of *types and names* à la Jäger [18]. Instead of *functions* in set theoretic sense, EM assumes the notion of *rule* or *algorithm* as basic. Similarly, a fundamental tenet is that a collection X always comes equipped with an *explicit presentation*, i.e. by specifying a *defining property* given by a. It follows that the membership predicate for stating that a given object x is a member of X is naturally interpreted by means of satisfaction or predicate application: $x \in X$ iff *a truly applies* to x, or x satisfies the defining property (presented by) a of X, a being termed a *name* of X (see [18]). This calls for a ground applicative structure $\mathcal{M}$ with a primitive application operation (a applies to x) and a truth predicate T which applies to elements of $\mathcal{M}$. It is a crucial point: T is not a metamathematical predicate in the standard sense, that applies to

This paper originates from the slides for the talk presented at the Jäger conference, Bern, December 12–13, 2013. We wish to thank the organizers for the nice hospitality. Thanks to an anonymous referee for comments and criticism.

A. Cantini (✉)
Dipartimento di Lettere e Filosofia, Università degli Studi di Firenze, Florence, Italy
e-mail: andrea.cantini@unifi.it

R. Kahle et al. (eds.), *Advances in Proof Theory*, Progress in Computer Science and Applied Logic 28, DOI 10.1007/978-3-319-29198-7_2

(the Gödel numbers of) formulas of a given formal language. Abstract truth applies in the present context to suitable objects—termed *propositions*—as distinguished from sentences: sentences may represent propositions, but it is in general open whether every sentence represents a proposition, and whether every proposition is explicitly represented by a sentence.

This leads to the motivation of the present paper: we like to investigate some connections between axiomatic theories of truth in the general setting explained above, and explicit mathematics. In this direction Jäger and his school have offered important contributions, which range from the field of (meta)predicativity to the field of feasible applicative systems, as witnessed e.g. by [20, 22] or [8].

As to the contents, we survey four different truth-theoretic extensions of a basic applicative theory **TON**, the core of explicit mathematics. Section 2 describes a compositional theory **CT** of propositions and truth, corresponding to Jäger's **EET**, the elementary theory of types and names, which is related to the system of arithmetical analysis, and ought to be compared with axiomatic theories of truth over Peano arithmetic, as investigated by Halbach [17]. We then consider two incompatible extensions of **CT**. Section 3 deals with an extension **AT**, where the very predicate of being a proposition does define a propositional function. **AT** justifies theories of types and names, where elementary types are closed under a weak form of power type operation. In Sect. 4 we describe a theory **PT** of propositions and truth, in which propositions and truth interact up to a certain point, and the collection of types is also closed under the so-called join axiom. Thus **PT** interprets **EETJ**, i.e. **EET** with the join principle. The conclusive Sect. 5 explores a strengthening of axiomatic abstract self-referential truth via generalized induction principles; this matches the theory **NEM** of **EM**, where a principle of name induction is assumed besides join and elementary comprehension.

Some results are only stated and proofs sketched. A detailed development is left to a subsequent paper. Nevertheless, there should be little somewhat novel points in comparison to the extant literature, e.g. Theorems 5–6 of Sect. 2.4, Theorem 38 of Sect. 4.3, and Theorem 46 in Sect. 5.2.

2 Truth and Types I

2.1 Abstract Truth Over Combinatory Structures

The picture to have in mind corresponds to a structure $\mathcal{M} = \langle M, \cdot_M, \mathcal{P}^M, \mathcal{T}^M \rangle$, such that:

1. M is an *expanded combinatory algebra* with binary total application $\cdot_M$ and distinguished elements representing (i) basic combinators; (ii) basic number theoretic constructors; (iii) logic constructors and predicates; as usual $|M|$ is the underlying non-empty set of M (*domain* of M);
2. M contains an isomorphic copy N_M of the standard natural numbers;

3. $\mathcal{P}^M$, $\mathcal{T}^M$ are distinguished subsets of $|M|$ containing (objects representing) *elementary propositions* and *true* elementary propositions.

   $\mathcal{T}^M$ acts as a *classifier* over $|M|$: there is a natural map $\mathcal{R}_M$ such that

$$a \in M \mapsto \mathcal{R}_M(a) := \{c \in M | \mathcal{T}^M(ac)\}. \tag{1}$$

$\mathcal{R}_M$ yields a connection with Jäger's approach to Explicit Mathematics. The set $\mathcal{R}_M(a)$ corresponds to *the type named by a in the sense of explicit mathematics.*

The collection of explicit types over $\mathcal{M}$ depends on the closure conditions on $\mathcal{P}^M$ and $\mathcal{T}^M$. And this is a reason to investigate abstract notions of truth with non-trivial closure properties. $\mathcal{P}^M$ may be absent and defined via $\mathcal{T}^M$.

2.2 Ground System: Language and Notations

The *basic first order language* $\mathcal{L}_T$ includes: predicate symbols $=$, P (proposition), T (true), N (natural number), the binary function symbol App (application), combinators K, S, the constant 0, successor SUC, predecessor PR, definition by cases on numbers $\mathsf{D_N}$, pairing PAIR, with projections LEFT, RIGHT; certain additional constants for representing predicate and logical constructors, namely: $\dot{=}$, $\dot{N}$, $\dot{P}$, $\dot{T}$, $\dot{\wedge}$, $\dot{\neg}$, $\dot{\forall}$. The constants $\dot{\vee}$, $\dot{\exists}$ and $\dot{\rightarrow}$ are also freely used (when needed), as being defined according to their classical definitions.[1]

$\mathcal{L}_P$ is the T-free sublanguage of $\mathcal{L}_T$ (i.e. T does not occur in $\mathcal{L}_P$-formulas); $\mathcal{L}_{op}$ is the P-free sublanguage of $\mathcal{L}_P$ (i.e. P does not occur in $\mathcal{L}_P$-formulas).

As usual, terms are inductively generated from variables and individual constants via application (for which we adopt the notation $ts := App(t, s)$). Formulas are inductively defined from atoms of the form $t = s$, $P(t)$, $T(t)$, $N(t)$ by means of sentential operators and quantifiers.

We further introduce *a map from formulas into terms*, preserving free variables:

$$A \mapsto [A].$$

E.g. $[t = s] := \dot{=}ts$, $[P(t)] := \dot{P}t$; $[T(t)] := \dot{T}t$; $[\forall xA] := \dot{\forall}(\lambda x.[A])$, etc.

Henceforth we also agree to use *infix notation* whenever we have constants representing binary operators or predicates, that is, we write $t \dot{\rightarrow} s$, $t \dot{\wedge} s$, etc. instead of the proper $\dot{\rightarrow} ts$, $\dot{\wedge} ts$, etc.

Abstraction terms can then be defined by lambda abstraction, i.e. if A is an arbitrary formula, $\{x|A\} := \lambda x.[A]$. We need some defined predicates:

[1] E.g. $\dot{\exists} f := \dot{\neg}(\dot{\forall}(\lambda x.\dot{\neg}(fx)))$; $\dot{\rightarrow}$ is assumed as primitive in Sect. 4.

1. $PF(f) := \forall x.P(fx), f$ is a *propositional function*;
2. $DF(f) := \forall x D(fx)$, where

$$D(a) := T(a) \vee T(\dot{\neg} a). \tag{2}$$

   $D(a)$ can be read as: a is *determinate* or *meaningful* (see [10, 13]), while $DF(f)$ means that f has *determinate truth values*.
3. $t =_e s := \forall u(T(tu) \leftrightarrow T(su))$; $=_e$ represents *extensional equality*.
4. $x\eta y := T(yx)$ is used for predicate application (or intensional membership).

2.3 Ground System: Applicative and Compositional Axioms

All systems we consider include a ground theory $\mathbf{TON}^-$ of total operations and numbers, i.e. equations for the standard combinators K and S, pairing, projections, closure axioms for the predicate N (natural numbers) and basic operations of successor SUC, predecessor PR, the constant 0, definition by cases $\mathsf{D_N}$ on N. For details, see [21, 22, 28].

In addition, we tacitly include in the ground system $\mathbf{TON}^-$ the following *independence conditions*: if b_0, b_1 (c, d) are distinct (arbitrary) constants among $\dot{=}$, $\dot{N}$, $\dot{P}$, $\dot{T}$, $\dot{\wedge}$, $\dot{\neg}$, $\dot{\forall}$:

$$b_0 \neq b_1, \tag{3}$$

$$cx = dy \rightarrow c = d \wedge x = y. \tag{4}$$

We also require some form of number theoretic induction; below we need a taxonomy of induction principles about numbers and propositional objects. We then state the basic axioms on propositions and we describe the core system.

Number-theoretic induction The schema $\mathcal{L}_T$-IND_N: if A is any arbitrary formula of $\mathcal{L}_T$, and $x + 1$ stands for the applicative term SUCx,

$$A(0) \wedge (\forall x \in N)(A(fx) \rightarrow A(f(x+1))) \rightarrow \forall x(N(x) \rightarrow A(fx)). \tag{5}$$

Besides $\mathcal{L}_T$-IND_N, we shall also consider other versions of the induction principle for N (in decreasing strength).

- The schema $\mathcal{L}_P$-IND_N: as $\mathcal{L}_T$-IND_N except that A is any formula of $\mathcal{L}_P$.
- The schema $\mathcal{L}_{op}$-IND_N: as $\mathcal{L}_T$-IND_N except that A is any formula of $\mathcal{L}_{op}$.
- The axiom PF-IND_N for propositional functions: add the hypothesis $PF(f)$ to (5) while replacing $A(-)$ by $T(f-)$.
- N-induction for determinate conditions D-IND_N: replace $PF(f)$ by $DF(f)$ in PF-IND_N.

Axioms for atomic propositions P-axioms

$$T(x) \to P(x), \tag{6}$$
$$P([x = y]) \wedge (T([x = y]) \leftrightarrow x = y), \tag{7}$$
$$P([N(x)]) \wedge (T([N(x)]) \leftrightarrow N(x)). \tag{8}$$

Axioms for classical compositional truth T-axioms

$$P(x) \to P(\dot{\neg} x) \wedge (T(\dot{\neg} x) \leftrightarrow \neg T(x)), \tag{9}$$
$$P(x) \wedge P(y) \to P(x \mathbin{\dot{\wedge}} y) \wedge (T(x \mathbin{\dot{\wedge}} y) \leftrightarrow T(x) \wedge T(y)), \tag{10}$$
$$\forall x P(fx) \to P(\dot{\forall} f) \wedge (T(\dot{\forall} f) \leftrightarrow \forall x T(fx)). \tag{11}$$

Axioms for strictness S-axioms:

$$P(\dot{\neg} x) \to P(x), \tag{12}$$
$$P(x \mathbin{\dot{\wedge}} y) \to P(x) \wedge P(y), \tag{13}$$
$$P(f) \to \forall x P(fx). \tag{14}$$

2.3.1 Simple Consequences of the Core System

Definition 1

1. **CT** is **TON**$^-$ with the truth axioms for atomic propositions, classical compositional truth, strictness and the schema of number theoretic induction $\mathcal{L}_T$-IND_N for arbitrary formulas of $\mathcal{L}_T$ (5).
2. **CT**$\lceil$ = **CT** with formula N-induction replaced by $\mathsf{PF\text{-}IND_N}$.

Notation. In general, given any formal theory SF, SF^- is the theory obtained from SF by omitting N-induction (of any sort).

Proposition 2 (provably in **CT**$^-$)

(i) Propositional objects are exactly the determinate ones in the sense of (2)*:*

$$\forall x(P(x) \leftrightarrow T(x) \vee T(\dot{\neg} x)). \tag{15}$$

Hence

$$\mathit{PF\text{-}IND_N} \leftrightarrow \mathit{D\text{-}IND_N}.$$

(ii) If A is a formula of $\mathcal{L}_{op}$ with free variables in the list $\vec{x} = x_1, \ldots, x_n$,

$$\forall x_1 \ldots \forall x_n (P([A(\vec{x})])). \tag{16}$$

(iii) Moreover, under the same assumption as (ii):

$$\forall x_1 \ldots \forall x_n (T([A(\vec{x})]) \leftrightarrow A(\vec{x})). \tag{17}$$

Proof Ad (i), (15): apply (6), (12) from right to left, and (9) from left to right.
Ad (ii)–(iii): easy induction on the definition of A, applying the closure conditions on the predicate P. □

Remark 1 (i) **CT** and **CT**$\lceil$ are neutral as to the internal status of P, i.e. we do not have in general $T([P(x)]) \vee T([\neg P(x)])$. In order to conclude $T([\neg P(x)])$ from $\neg T([P(x)])$, we need that $P(x)$ *is a proposition*: this is an axiom of the next system to be considered.
(ii) If we interpret $P(x)$ trivially, i.e. we assume that *everything is a proposition*, all axioms involving P go through except the one for negation: for then we should have to postulate $T(a) \vee T(\dot{\neg} a)$, which leads to inconsistency.

***P*-induction** A natural assumption on the set of propositions is to assume that it is inductively generated according to clauses embodied by the axioms on atomic propositions and compositional truth.

Definition 3 Let the corresponding positive elementary operator be given by the formula

$$\begin{aligned} \mathcal{C}(u, X) \quad \Leftrightarrow \quad & \exists x \exists y [(u = [x = y] \vee u = [N(x)] \vee \\ & \vee (u = \dot{\neg} x \wedge X(x)) \vee (u = (x \dot{\wedge} y) \wedge X(x) \wedge X(y)) \vee \\ & \vee (u = \dot{\forall} x \wedge \forall z X(xz)))]. \end{aligned}$$

(i) Then $\mathcal{L}_T$-IND_P is the schema of induction on propositions:

$$\forall x (\mathcal{C}(x, B) \rightarrow B(x)) \rightarrow \forall x (P(x) \rightarrow B(x)) \tag{18}$$

where $B(x)$ is an arbitrary formula of $\mathcal{L}_T$.
(ii) If we restrict B to be a formula of $\mathcal{L}_P$ (respectively of $\mathcal{L}_{op}$), we have the corresponding schema $\mathcal{L}_P$-IND_P($\mathcal{L}_{op}$-IND_P). Similarly, PF-IND_P(D-IND_P) is the P-induction axiom restricted to propositional (determinate) functions (choose $B(x) := T(fx)$ and assume fx to be a proposition).

Lemma 4 *The strictness axioms for propositions become provable in* **CT**$^-$ *with* $\mathcal{L}_P$-IND_P *restricted to formulas which are positive in P.*

(Simply choose $B(x) := \mathcal{C}(x, P)$ in $\mathcal{L}_P$-IND_P and apply the independence axioms (3), (4)).

2.4 Conservativity and Upper Bound

First of all, a preliminary remark. We assume that the reader is acquainted with the relationship between formal systems: $\mathcal{S}$ *is proof-theoretically reducible to* $\mathcal{R}$, in short $\mathcal{S} \leq \mathcal{R}$ [12, 13]. It is understood that the relation holds modulo some fixed class of formulas in the common language and a fixed metatheory U, where the reduction proof can be carried out. Usually the reduction consists of an effective method for transforming proofs in $\mathcal{S}$ into proofs in $\mathcal{R}$, which is shown to converge in U and to preserve formulas of the given class Φ. As standard choice, U is primitive recursive arithmetic **PRA**, while Φ includes at least the formulas expressing the totality of operations from N to N. We also say that $\mathcal{S}$ is *proof-theoretically equivalent* to $\mathcal{R}$ (in short $\mathcal{S} \equiv \mathcal{R}$) iff $\mathcal{S} \leq \mathcal{R}$ and $\mathcal{R} \leq \mathcal{S}$; $\leq$ is so defined that it is reflexive and transitive, and hence $\equiv$ is an equivalence.

We further assume acquaintance with the relation: $\mathcal{S}$ *is relatively interpretable in* $\mathcal{R}$.

With this in mind we pass to consider the following problem: is $\mathbf{CT}\lceil$ conservative with respect to **TON**?

Theorem 5 $\mathbf{CT}\lceil$ + $\mathcal{L}_T$-IND_P *is proof-theoretically reducible to* **TON** *(and actually conservative over* **TON***).*

Proof This can be shown by interpreting $\mathbf{CT}\lceil$ in the least fixed point model of the abstract Kripke-Feferman theory, which is mirrored axiomatically by the system $\mathbf{KF}\lceil+\mathbf{GID}$ and whose upper bound is **TON** (see [3, 6]). To this aim, we assume that the reader has skipped just a moment to Sect. 5 where the relevant interpreting theory $\mathbf{KF}\lceil+\mathbf{GID}$ is described.

First of all, the applicative language of $\mathbf{CT}\lceil$ may be regarded as a sublanguage of the language of $\mathbf{KF}\lceil+\mathbf{GID}$. Indeed, let * be the map which preserves application, it is the identity on the **TON**-constants (combinators, number-theoretic operations) and is defined as follows on the special constants[2]:

$$
\begin{aligned}
\mathsf{eq}^* &= \lambda x \lambda y.\langle 1, \langle x, y\rangle\rangle,\\
\mathsf{nat}^* &= \lambda x.\langle 2, x\rangle,\\
\dot{\wedge}^* &= \lambda x \lambda y.\langle 3, \langle x, y\rangle\rangle,\\
\dot{\neg}^* &= \lambda x.\langle 4, x\rangle,\\
\dot{\forall}^* &= \lambda x.\langle 5, x\rangle.
\end{aligned}
$$

Choose P^* by the fixed point theorem for predicates in $\mathbf{KF}\lceil$, so that $\forall x(P^*(x) \leftrightarrow \mathcal{C}(x, P^*))$ where $\mathcal{C}(x, -)$ is the positive elementary operator of (18). We argue in in $\mathbf{KF}\lceil+\mathbf{GID}$ *by generalized induction on* P^* and show that every proposition in the sense of P^* has a determinate truth value, i.e.

[2] We use the standard abbreviations $\langle t, s\rangle := \mathsf{PAIR}ts$; $(t)_0 := \mathsf{LEFT}t$, $(t)_1 := \mathsf{RIGHT}t$. Below 1, 2, … stand for the corresponding numerals.

$$\forall x(P^*(x) \to T(x) \vee T(\dot{\neg}x)). \tag{19}$$

We then extend the star map to a translation $A \mapsto A^*$ of the language of **CT**⌈ into the language of **KF**⌈+**GID**, such that

$$(P(x))^* \quad := \quad P^*(x), \tag{20}$$

$$(T(x))^* \quad := \quad T(x) \wedge P^*(x). \tag{21}$$

The essential step is (19), which requires a special instance of **GID** which is positive in T. Also observe that, again by (19), the axiom of N-induction for *propositional functions* is sent onto an instance of N-induction for determinate conditions. Hence the *-translations of all theorems of **CT**⌈ are transformed into theorems of **KF**⌈+**GID**. Moreover, the *-translation of $\mathcal{L}_T$-IND_P becomes provable by **GID** and every theorem of **CT**⌈ in the applicative part of the language is also provable in **KF**⌈+**GID**. But **KF**⌈+**GID** is proof-theoretically reducible to and conservative over **TON** by [3, 6] whence the claim. □

Remark 2 (i) We *cannot prima facie directly embed* **CT**⌈ into **KF**⌈ by identifying $P(x)$ with $T(x) \vee T(\dot{\neg}x)$ because of the strictness conditions; this is the reason we have to pass through the inductive definition of P in **KF**⌈+**GID**.

(ii) On the surface, one is tempted to consider **CT** as an abstraction from and an analogue to Halbach's system **CT**[**PA**] of compositional truth over **PA** (see [17]). And it is known that **CT**[**PA**] is conservative over **PA** [24]. However, the analogy is only superficial, as the notion of sentence in the case of **CT**[**PA**] is arithmetically fixed and does not depend on an inductive definition.

Theorem 6 **CT** + $\mathcal{L}_T$-IND_P *and arithmetical analysis* **ACA** *are mutually interpretable.*

Proof

(i) Lower bound: it follows from Theorem 7 (see below).

(ii) Upper bound. First of all, we fix an arithmetization of the open term model TER for **TON**, induced by the underlying expanded combinatory logic; this can be done within theories of strength at most **PA** (for details see e.g. [21, 28]). Then we devise *an arithmetical interpretation* $\mathcal{P}^{TM}$ for the predicate P by means of a *recursively enumerable* derivability relation $\vdash$. The axioms of $\vdash$ have the form

$$\frac{}{\vdash [t = s]} \qquad \frac{}{\vdash [N(t)]}$$

for the basic atomic formulas with $=$ and N, where t, s range over elements of TM. The inference rules corresponding to negation, conjunction and universal quantification have the form:

$$\frac{\vdash t}{\vdash \dot{\neg}t} \qquad \frac{\vdash t \quad \vdash s}{\vdash (t \dot{\wedge} s)}$$

The clause for $\forall$ is rephrased as a *finitary* inference:

$$\frac{\vdash\ tx}{\vdash\ \dot{\forall} t}$$

provided x is not free in t.

It is then easy to check that the derivability relation is *closed under substitution*, that is, for arbitrary terms t, s:

$$\vdash t\ \Rightarrow \vdash t[x := s]. \tag{22}$$

Each statement $\vdash t$ can be naturally attached a number k such that $\vdash^k t$ means that the deduction tree for t has depth k. Then define

$$t \in \mathcal{P}^{TM} :\Leftrightarrow \exists k(\vdash^k\ t\ \wedge t \in TM). \tag{23}$$

We can verify:

(i) $\mathcal{P}^{TM}$ satisfies the P-axioms in **ACA**.
(ii) $\mathcal{P}^{TM}$ satisfies (the interpretation of) $\mathcal{L}_T$-IND_P in the open term model.

As to (ii), it amounts to check by number theoretic induction on k and (22) that $\forall k \forall t(\vdash^k\ t \rightarrow B^{TM}(t))$ holds under the assumption that the subset defined by $B(x)$ in TM is closed under the clauses generating propositions. Finally, we can define an interpretation for T by simulating the canonical definition of the truth predicate for arithmetic in the subsystem **ACA** of second order arithmetic, based upon arithmetical comprehension. □

2.5 Elementary Types

The language $\mathcal{L}_\nu$ for the elementary theory **EET** of types and names consists of:

(i) predicate constants $=$ (identity predicate for individuals), $=_t$ (identity predicate for types), $\in$ (membership), $\mathcal{R}$ (naming relation), N (natural numbers);
(ii) besides the usual individual constants for the extended combinatory algebras, one has naming operations for types: nat, id, co, int, inv, dom;
(iii) countably many individual variables $(x, y, z \ldots)$ and type variables $X, Y, Z \ldots$.

EET obviously includes **TON**$^-$, independence conditions for constructors[3] and the number theoretic induction schema. **EET**$\lceil$ is the corresponding subsystem with *the axiom of N-induction for types*:

[3]Among them nat, id, co, int, inv, dom; we leave the obvious statement to the reader, in analogy to (3), (4).

$$0 \in X \wedge (\forall x \in N)(x \in X \to (x+1) \in X) \to \forall x(N(x) \to x \in X). \qquad (24)$$

EET: representation axioms and extensionality

R1 $\forall X \exists y \mathcal{R}(y, X)$;
R2 $\mathcal{R}(a, X) \wedge \mathcal{R}(a, Y) \to X =_t Y$;
Ext $\forall x(x \in X \leftrightarrow x \in Y) \to X =_t Y$.

In sum: there exists a partial surjection $\mathcal{R}$ from the universe V onto *TYPE*:

$$\mathcal{R} : W \twoheadrightarrow TYPE,$$

for some $W \subseteq V$.

EET: existence of elementary types

nat $\exists X(\mathcal{R}(\mathsf{nat}, X) \wedge \forall x(x \in X \leftrightarrow N(x)))$;
id $\exists X(\mathcal{R}(\mathsf{id}, X) \wedge \forall x(x \in X \leftrightarrow \exists y(x = \mathsf{PAIR}yy)))$;
co $\mathcal{R}(y, Y) \to \exists X(\mathcal{R}(\mathsf{co}y, X) \wedge \forall x(x \in X \leftrightarrow \neg x \in Y))$;
int $\mathcal{R}(y, Y) \wedge \mathcal{R}(z, Z) \to \exists X(\mathcal{R}(\mathsf{int}yz, X) \wedge \forall x(x \in X \leftrightarrow x \in Y \wedge x \in Z))$;
inv $\mathcal{R}(y, Y) \to \exists X(\mathcal{R}(\mathsf{inv}fy, X) \wedge \forall x(x \in X \leftrightarrow fx \in Y))$;
dom $\mathcal{R}(y, Y) \to \exists X(\mathcal{R}(\mathsf{dom}y, X) \wedge \forall x(x \in X \leftrightarrow \exists v(\mathsf{PAIR}xv \in Y)))$.

CT and EET

Theorem 7

(i) **EET**$\lceil$ *is interpretable into* **CT**$\lceil$.
(ii) **EET** *is interpretable into* **CT**.

Proof We define a suitable embedding. □

Definition 8 The translation $A \mapsto A^*$.

(i) * preserves application, it is the identity transform on **TON**-constants (combinators, number-theoretic operations) and variables.[4] Moreover the *-transform of the special constants is defined as follows:

1. $\mathsf{id}^* = \lambda x[\exists y(x = \langle y, y\rangle)]$,
2. $\mathsf{nat}^* = \lambda x[N(x)]$,
3. $\mathsf{dom}^* = \lambda y \lambda u \dot{\exists} v(y(\langle u, v\rangle))$,
4. $\mathsf{int}^* = \lambda x \lambda y \lambda u.((xu) \dot{\wedge} (yu))$,
5. $\mathsf{co}^* = \lambda x \lambda u.\dot{\neg}(xu)$,
6. $\mathsf{inv}^* = \lambda y \lambda f \lambda x.y(fx)$.

[4]This makes sense, since we can identify individual variables of **EET** with **CT**-variables with odd indices, and type variables of **EET** with **CT**-variables with even indices.

(ii) Atomic formulas:

$$\mathcal{R}(a, X)^* := (a =_e X) \wedge PF(a) \wedge PF(X), \quad (25)$$
$$(X =_t Y)^* := X =_e Y, \quad (26)$$
$$(a \in X)^* := T(X(a)), \quad (27)$$
$$(t = s)^* := (t^* = s^*), \quad (28)$$
$$N(t)^* := N(t^*). \quad (29)$$

(For $=_e$, $PF(x)$ in (25) above, see Sect. 2.2.)

(iii) $A \mapsto A^*$ is extended to the non-atomic formulas by requiring that it preserves sentential connectives, individual quantification. In addition:

$$(\forall XA)^* := \forall x(PF(x) \rightarrow A^*[X := x]), \quad (30)$$
$$(\exists XA)^* := \exists x(PF(x) \wedge A^*[X := x]). \quad (31)$$

It is enough to check that the translation of the axioms of **EET**$\lceil$ is provable in **CT**$\lceil$.

(i) R1–R2, Ext: the translations are tautologies
(ii) nat-id-co-int-inv-dom: apply the corresponding axioms for atomic propositions and classical compositional truth. For instance, in the case of inverse image, if y is a propositional function, then $\mathsf{inv}^* fy := \lambda x.y(fx)$ is a propositional function of x given that y is, and it satifies the *-transform of

$$\exists X(\mathcal{R}(\mathsf{inv}fy, X) \wedge \forall x(x \in X \leftrightarrow fx \in Y)).$$

3 Truth and Types II

3.1 *Strengthening* CT*: The Theory* AT

Does the notion of proposition define a propositional function, i.e. is $\lambda x.[P(x)]$ a propositional function? We consider theories which yield a positive answer.

Definition 9

$$P([P(x)]) \wedge (T([P(x)]) \leftrightarrow P(x)). \quad (32)$$

1. **AT** := **CT** + (32),
2. **AT**$\lceil$:= **CT**$\lceil$ + (32).

Lemma 10

(i) *Let S be such that $S = [P(S)]$. Then* $\mathbf{AT}^- \vdash P(S) \wedge T(S)$.
(ii) *Let L be such that $L = [\neg P(L)]$. Then* $\mathbf{AT}^- \vdash P(L) \wedge T(\dot{\neg} L)$.
(iii) $\mathbf{AT}^-$ *proves:*

$$\forall x(T([P(x)]) \vee T([\neg P(x)])), \tag{33}$$
$$\forall x(T(x) \rightarrow T([P(x)])), \tag{34}$$
$$\forall x T([P([P(x)])]). \tag{35}$$

Proof

(i): by the axiom (32), $P([P(S)])$; hence $P(S)$, so $T([P(S)]))$, hence $T(S)$.
(ii): by (32) and (12) we obtain $P(L)$. Were $T(L)$, then $T([\neg P(L)])$ and hence $\neg T([P(L)])$, i.e. $\neg P(L)$: contradiction. Hence $\neg T(L)$, i.e. $T(\dot{\neg} L)$.
(iii): by axioms (6), (9), (32). □

Informally, (i)–(ii) above can be rendered as: there is a true proposition saying "I am a proposition", there is a false proposition saying "I am not a proposition".

Lemma 11 *Proposition 2 holds for* $\mathbf{AT}^-$*: the Tarski schema (see* (16)*,* (17)*) holds for every formula A of $\mathcal{L}_P$. In particular, every formula A of $\mathcal{L}_P$ with free variables in the list $\vec{x} = x_1, \ldots, x_n$ defines a propositional function:*

$$\forall x_1 \ldots \forall x_n (P([A(\vec{x})])). \tag{36}$$

Remark 3 Under the same label **AT** a related system was introduced in [5]; the original system included $\mathbf{CT}^-$ with (37)–(39):

$$P([T(x)]) \leftrightarrow P(x), \tag{37}$$
$$T([T(x)]) \leftrightarrow T(x), \tag{38}$$
$$T([\neg P(x)]). \tag{39}$$

It is consistent to expand the present **AT** with (37), (38): simply put $[T(x)] := x$. Also, the old system proved $P([P(x)])$ but *it is incompatible* with the present one.

3.2 Generating an AT*-Model: Propositions*

As usual, if $\mathcal{M}$ is a structure for a given language $\mathcal{L}$ (among those described), $\mathcal{L}^M$ is $\mathcal{L}$ expanded with distinct constants for distinct elements of the domain $|M|$. For the sake of simplicity, we keep using the same notation a for an element of $|M|$ and the corresponding constant. If t is a closed term of the expanded language, t^M is the unique element of $|M|$ denoted by t in $\mathcal{M}$.

Definition 12 We recursively define the collection of propositional objects uniformly in any given expanded combinatory algebra.

- Initial clause: if $A := (a = b),\ Na,\ Pa,$

$$\mathcal{P}_0^M = \{[A]^M \mid a,\ b \in M\}.$$

- Successor clause:

$$\begin{aligned}\mathcal{P}_{\alpha+1}^M \quad = \quad & \mathcal{P}_\alpha^M \cup \{(\dot{\neg} a)^M | a \in \mathcal{P}_\alpha^M\} \cup \\ & \cup \{(b \mathbin{\dot{\wedge}} c)^M | b \in \mathcal{P}_\alpha^M,\ c \in \mathcal{P}_\alpha^M\} \cup \\ & \cup \{(\dot{\forall} f)^M | \text{for all } c \text{ in } |M|, fc^M \in \mathcal{P}_\alpha^M\}.\end{aligned}$$

- If λ is a limit,

$$\mathcal{P}_\lambda^M = \bigcup \{\mathcal{P}_\beta^M \mid \beta < \lambda\}.$$

- Let

$$\mathcal{P}^M = \bigcup \{\mathcal{P}_\beta^M \mid \beta < card(M)^+\}.$$

Lemma 13

(i) For every α, β,

$$\alpha \leq \beta \Rightarrow \mathcal{P}_\alpha^M \subseteq \mathcal{P}_\beta^M. \tag{40}$$

(ii) If $a,\ b \in |M|$ and $A := (a = b),\ Na,\ Pa,$ then $([A])^M \in \mathcal{P}^M$.
(iii) Moreover:

$$a \in \mathcal{P}^M \Leftrightarrow (\dot{\neg} a)^M \in \mathcal{P}^M \tag{41}$$

$$(a \mathbin{\dot{\wedge}} b)^M \in \mathcal{P}^M \Leftrightarrow a \in \mathcal{P}^M \wedge b \in \mathcal{P}^M \tag{42}$$

$$\forall c \in |M| (ac)^M \in \mathcal{P}^M \Leftrightarrow (\dot{\forall} a)^M \in \mathcal{P}^M. \tag{43}$$

Proof Trivial transfinite induction on ordinals using the closure properties embodied in the definitions of $\mathcal{P}^M$. □

Incidentally the construction ensures that there are models of an auxiliary theory of operations and propositions $\mathbf{TON}_P$ formalized in the T-free language.

Definition 14 $\mathbf{TON}_P$ includes, besides the applicative axioms of **TON**:

1. the axioms concerning the predicate P, i.e. the T-free part of (7)–(14) and (32)[5];
2. N-induction and P-induction schemata restricted to $\mathcal{L}_P$-formulas.

[5]Concerning (32), we are thus left only with $\forall x P([P(x)])$.

Lemma 15 **TON**$_P$ *is interpretable into* **TON**.

The lemma holds since the open term model for **TON**$_P$ can be formalized in **TON** (for the proof, see Theorem 6).

Definition 16 If M is a model of **TON**$^-$, a structure $\langle M, \mathcal{P}^M\rangle$ for the language $\mathcal{L}_P$ which is a model of **TON**$_P$ is N-standard iff the denotation N^M of N is isomorphic to the structure of natural numbers; if, in addition, $\mathcal{P}^M$ is the least fixed point of the positive elementary operator $\mathcal{C}_p(x, -)$[6] inductively generating the notion of proposition over M according to Sect. 3.2, $\langle M, \mathcal{P}^M\rangle$ is called N, P-standard.

Trivially:

Lemma 17 *If M is N-standard and $\mathcal{P}^M$ is defined as above (see Definition 12), then $\langle M, \mathcal{P}^M\rangle$ is N, P-standard and $\langle M, \mathcal{P}^M\rangle \models$* **TON**$_P$.

3.3 Generating an AT-Model: Truth

A N, P-standard model of **TON**$_P$ can be expanded to a model of **AT**. Indeed, we produce a sequence $\{\mathcal{T}^M_\alpha\}$ approximating the truth set, uniformly in $\mathcal{P}^M$:

- Initial clause:

$$
\begin{aligned}
\mathcal{T}^M_0 \quad = \quad & \{[a=b]^M | M \models (a=b)\} \cup \\
& \cup \{[\neg a = b]^M | M \models \neg(a=b)\} \cup \\
& \cup \{[N(a)]^M | M \models N(a)\} \cup \\
& \cup \{[\neg N(a)]^M | M \models \neg N(a)\} \cup \\
& \cup \{[P(a)]^M | a \in \mathcal{P}^M\} \cup \\
& \cup \{[\neg P(a)]^M | a \notin \mathcal{P}^M\}.
\end{aligned}
$$

- Successor clause:

$$
\begin{aligned}
\mathcal{T}^M_{\alpha+1} \quad = \quad & \mathcal{T}^M_\alpha \cup \{(\dot\neg\dot\neg b)^M | b \in \mathcal{T}^M_\alpha\} \cup \\
& \cup \{(b \dot\wedge c)^M | b \in \mathcal{T}^M_\alpha \wedge c \in \mathcal{T}^M_\alpha\} \cup \\
& \cup \{(\dot\neg(b \dot\wedge c))^M | ((\dot\neg b)^M \in \mathcal{T}^M_\alpha \wedge (\dot\neg c)^M \in \mathcal{P}^M) \vee \\
& \vee ((\dot\neg c)^M \in \mathcal{T}^M_\alpha \wedge (\dot\neg b)^M \in \mathcal{P}^M)\} \cup \\
& \cup \{(\dot\forall f)^M | \text{ for all } c \text{ in } |M| \ (fc)^M \in \mathcal{T}^M_\alpha\} \cup \\
& \cup \{(\dot\neg\dot\forall f)^M | \text{ for some } c \text{ in } |M| \ (\dot\neg fc)^M \in \mathcal{T}^M_\alpha \wedge \\
& \wedge \ \text{ for all } c \text{ in } |M| (\dot\neg fc)^M \in \mathcal{P}^M)\}.
\end{aligned}
$$

[6]NB: this operator is distinct from $\mathcal{C}(x, -)$ of Remark 1 and schema (18), since it embodies the initial condition ensuring $P[(P(x))]$.

- If λ is a limit,

$$\mathcal{T}_{\lambda}^{M} = \bigcup \{\mathcal{T}_{\beta}^{M} \mid \beta < \lambda\}.$$

Lastly, define

$$\mathcal{T}^{M} = \bigcup \{\mathcal{T}_{\beta}^{M} \mid \beta < card(M)^{+}\}.$$

Remark 4 The internal truth value of $P(a)$ at stage 0 is determined *in an impredicative way*, i.e. by assuming the set of propositional objects as completed.

Lemma 18 *If $a \in |M|$, then for every α, β*

$$\alpha \leq \beta \quad \Rightarrow \quad \mathcal{T}_{\alpha}^{M} \subseteq \mathcal{T}_{\beta}^{M}, \tag{44}$$

$$a \in \mathcal{T}_{\alpha}^{M} \quad \Rightarrow \quad a \in \mathcal{P}^{M}, \tag{45}$$

$$a \in \mathcal{P}_{\alpha}^{M} \quad \Rightarrow \quad a \in \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} a)^{M} \in \mathcal{T}_{\alpha}^{M}, \tag{46}$$

$$a \in \mathcal{P}_{\alpha}^{M} \quad \Rightarrow \quad a \notin \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} a)^{M} \notin \mathcal{T}_{\alpha}^{M}. \tag{47}$$

Hence completeness and consistency hold:

$$a \in \mathcal{P}^{M} \Rightarrow (a \in \mathcal{T}^{M} \;\vee\; (\dot{\neg} a)^{M} \in \mathcal{T}^{M}) \wedge (a \notin \mathcal{T}^{M} \;\vee\; (\dot{\neg} a)^{M} \notin \mathcal{T}^{M}). \tag{48}$$

Proof Transfinite induction on ordinals using the closure properties embodied in the definitions of $\mathcal{T}^{M}$, $\mathcal{P}^{M}$. Let us only deal with (46).

Ad (46): assume $a \in \mathcal{P}_{0}^{M}$: if $a = [P(b)]^{M}$, then either $b \in \mathcal{P}^{M}$ and hence $[P(b)]^{M} \in \mathcal{T}_{0}^{M}$, or else $b \notin \mathcal{P}^{M}$ and hence by definition $[\neg P(b)]^{M} \in \mathcal{T}_{0}^{M}$. The other atomic cases are immediate.

Let $a = (\dot{\neg} b)^{M} \in \mathcal{P}_{\alpha+1}^{M}$. Then $b \in \mathcal{P}_{\alpha}^{M}$ and by IH, $b \in \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha}^{M}$. Hence by (44) and closure conditions on truth, $(\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha+1}^{M} \vee (\dot{\neg}\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha+1}^{M}$.

Let $a = (b \mathbin{\dot{\wedge}} c)^{M} \in \mathcal{P}_{\alpha+1}^{M}$, i.e. $b \in \mathcal{P}_{\alpha}^{M}$ and $c \in \mathcal{P}_{\alpha}^{M}$. Then by IH,

$$b \in \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha}^{M}, \tag{49}$$

$$c \in \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} c)^{M} \in \mathcal{T}_{\alpha}^{M}. \tag{50}$$

If $b \in \mathcal{T}_{\alpha}^{M}$ and $c \in \mathcal{T}_{\alpha}^{M}$, then by the closure conditions on truth, $(b \mathbin{\dot{\wedge}} c)^{M} \in \mathcal{T}_{\alpha+1}^{M}$. If $c \in \mathcal{T}_{\alpha}^{M}$ but $(\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha}^{M}$, then by (45) we have $c \in \mathcal{P}^{M}$ and $(\dot{\neg} b)^{M} \in \mathcal{T}_{\alpha}^{M}$, whence by the closure conditions on truth, $(\dot{\neg}(b \mathbin{\dot{\wedge}} c))^{M} \in \mathcal{T}_{\alpha+1}^{M}$. The symmetric case is similar.

Let $a = (\dot{\forall} f)^{M} \in \mathcal{P}_{\alpha+1}^{M}$. Then for all $c \in |M|$, $fc \in \mathcal{P}_{\alpha}^{M}$, whence by IH, for all $c \in |M|$:

$$fc \in \mathcal{T}_{\alpha}^{M} \vee (\dot{\neg} fc)^{M} \in \mathcal{T}_{\alpha}^{M} \tag{51}$$

which implies either $(\dot{\forall} f)^{M} \in \mathcal{T}_{\alpha+1}^{M}$ or, for some $c \in |M|$, $(\dot{\neg} fc)^{M} \in \mathcal{T}_{\alpha}^{M}$, i.e. since $(\dot{\forall} f)^{M} \in \mathcal{P}^{M}$, $(\dot{\forall} f)^{M} \in \mathcal{T}_{\alpha+1}^{M} \vee (\dot{\neg}\dot{\forall} f)^{M} \in \mathcal{T}_{\alpha+1}^{M}$.

If $a \in \mathcal{P}^M_\lambda$ with λ limit, apply IH and (40). □

Hence:

Proposition 19 *If* $\langle M, \mathcal{P}^M \rangle \models$ $\mathbf{TON}_P$ *is a* N, P*-standard model of* $\mathbf{TON}_P$*, then*

$$\langle M, \mathcal{P}^M, \mathcal{T}^M \rangle \models \mathbf{AT}. \tag{52}$$

3.4 On the Strength of AT

Theorem 20 **AT** *is proof theoretically equivalent to* **ACA**.

Proof

(i) Lower bound: obvious since **AT** extends **CT**.
(ii) Upper bound: by a straightforward extension of the proof of Theorem 6. Indeed, we first slightly modify the open term model TER for **TON** by adding to the derivability relation $\vdash$ a clause

$$\frac{}{\vdash [P(t)]}$$

The meaning of Pt becomes as before $\vdash^k t$, for some natural number k. Then we inductively define a sequence $\{\mathcal{T}^M_k\}$ satisfying the same closure conditions as those given in Sect. 3.3. The truth predicate is then interpreted as the union over the corresponding countable sequence and the finitary interpretation of P allows us to replace *arbitrary ordinals* by finite ones, to the extent that the whole construction can be carried out in **ACA** by a suitable non-arithmetical instance of number-theoretic induction. □

Conjecture 1 **AT**$\lceil$ *is proof-theoretically reducible to* **PA**.

3.5 Adding a Weak Power Type Operation

3.5.1 Axioms for Explicit Types: Weak Power

$\mathbf{EET}^\pi$ ($\mathbf{EET}^\pi\lceil$) is the extension of **EET** (**EET**$\lceil$) with the axiom (U)

$$\exists X(\mathcal{R}(\mathsf{cl}, X) \wedge \forall x(x \in X \leftrightarrow \exists Y(\mathcal{R}(x, Y))).$$

cl is the object representing the type of all names; intensionally (U) states the existence of *the type of all types*.

cl exists if the axiom *everything is a name* is assumed (and hence $\mathcal{R}$ is a surjection defined on V).

Proposition 21 (Uniform weak power type axiom [9]) *There exists a term π such that* $\mathbf{EET}^{\pi}\lceil$ *proves:*

$$\begin{aligned}&\mathcal{R}(x, X) \to \exists Y(\mathcal{R}(\pi x, Y) \wedge \\ &\wedge \forall Z(Z \subseteq X \to \exists v(v \in Y \wedge \mathcal{R}(v, Z)) \\ &\wedge \forall v(v \in Y \to \exists Z(Z \subseteq X \wedge \mathcal{R}(v, Z)))))\end{aligned} \tag{53}$$

For more information about power types in explicit mathematics, see also [4, 19].

3.5.2 Embedding $\mathbf{EET}^{\pi}$ into AT

We extend the translation $A \mapsto A^*$ of Definition 8 with an additional clause:

$$\mathsf{cl}^* = \lambda u[PF(u)].$$

Theorem 22 $\mathbf{EET}^{\pi} \vdash A \Rightarrow \mathbf{AT} \vdash A^*$ *and the same holds for the pair* $\mathbf{EET}^{\pi}\lceil$ *and* $\mathbf{AT}\lceil$.

Proof By Theorem 7 it is enough to verify the *-transform of the axiom (U)

$$\exists X(\mathcal{R}(\mathsf{cl}, X) \wedge \forall x(x \in X \leftrightarrow \exists Y(\mathcal{R}(x, Y))$$

holds. Therefore we check

$$\begin{aligned}&PF(\lambda x.[PF(x)]),\\ &\forall u(T(\lambda x[PF(x)])u \leftrightarrow \exists Y(PF(Y) \wedge PF(u) \wedge u =_e Y)).\end{aligned}$$

The *-transform of the second condition is trivial (choose $Y = u$). As to the first one, we have by the first part of axiom (32), the definition of propositional function, β-conversion and (11):

$$\begin{aligned}\Rightarrow&\quad \forall v P([P(v)])\\ \Rightarrow&\quad \forall u \forall x P([P(ux)])\\ \Rightarrow&\quad \forall u P([\forall x P(ux)])\\ \Rightarrow&\quad \forall u P([PF(u)])\\ \Rightarrow&\quad \forall u P((\lambda x.[PF(x)])u)\\ \Rightarrow&\quad PF(\lambda x.[PF(x)]).\end{aligned}$$

□

Theorem 23 $\mathbf{EET}^{\pi}\lceil$ *(*$\mathbf{EET}^{\pi}$*) is proof-theoretically reducible to* **PA** *(*$\mathbf{ACA}$*).*

This is known to be true by [16]. It follows by Theorem 22 if Conjecture 1 is true.

4 Truth and Types III

4.1 Strengthening CT*: The System* PT

We introduce a theory of propositions and truth, where the constant $\dot{\rightarrow}$ is a primitive symbol.[7]

Definition 24 (i) **PT**:=**CT** + (54) + (55) + (56), where

$$P(x) \quad \leftrightarrow \quad P([P(x)]), \tag{54}$$

$$P(a \dot{\rightarrow} b) \quad \leftrightarrow \quad P(a) \wedge (T(a) \rightarrow P(b)), \tag{55}$$

$$P(a \dot{\rightarrow} b) \quad \rightarrow \quad (T(a \dot{\rightarrow} b) \leftrightarrow (T(a) \rightarrow T(b))). \tag{56}$$

(ii) **PT**$\lceil$ is **PT** with number-theoretic induction *restricted to propositional functions.*
(iii) $a \odot b := a \dot{\wedge} (a \dot{\rightarrow} b)$.[8]

Remark 5 Essentially the same system (also labelled **PT**) was presented to the Russell conference in München in 2001 (see [5]) in the context of the discussion of a Russellian paradox about truth and propositions. If we compare it with systems available in the literature, **PT** and its variants are closely related to Aczel's theory of Frege structures (see [1]). Aczel's axioms do not include (54) and (56) while (55) is only stated from right to left and also the strictness conditions for propositions are not included. **PT** can be regarded as abstract version of *Feferman's theory* **DT** *of determinate truth* (see [10, 13], p. 318) over Peano Arithmetic. Here **PA** is replaced by the applicative theory **TON**, Feferman's determinatess predicate D is interpreted as P and assumed as primitive.[9]

Lemma 25 (provably in **PT**$^-$)

(i) Assume that a is a proposition and that b is a proposition provided a is true. Then

$$P(a \odot b) \wedge (T(a \odot b) \leftrightarrow T(a) \wedge T(b))). \tag{57}$$

(ii) Assume that f is a family of propositional functions indexed by the propositional function a. Choose $t(u) := [u = (u_0, u_1)]$, $s(u) := (au_0) \odot (fu_0)u_1$, *and define* $\mathsf{j}(a, f) := \lambda u.(t(u) \dot{\wedge} s(u))$. *Then* $\mathsf{j}(a, f)$ *is a propositional function, such that*

$$T(\mathsf{j}(a, f)u) \leftrightarrow T(au_0) \wedge T((fu_0)u_1). \tag{58}$$

[7]Hence (3), (4) are expanded so as to include $\dot{\rightarrow}$.

[8]This is a sequential conjunction introduced by Aczel in [1].

[9]Indeed Feferman [10], noting that Aczel's approach is based on λ-calculus which allows for more general interpretations, adds that "further work on systems like **DT** might usefully incorporate similar features."

Proof (i): apply (10), (55), (56). As to (ii), simply apply (i). □

Remark 6 **PT**$^-$ can be conservatively extended by

$$P([T(x)]) \leftrightarrow P(x), \quad (59)$$
$$T([T(x)]) \leftrightarrow T(x), \quad (60)$$
$$T([\neg T(x)]) \leftrightarrow T(\dot{\neg} x). \quad (61)$$

Simply define $[Tt] := t$.

The negation axioms for P and T are redundant once we assume an implication operator as primitive. Define $\dot{\neg} a := (a \dot{\rightarrow} \bot)$, where $\bot := [K = S]$. Then since **PT**$^-$ without axioms on $\dot{\neg}$ derives $P(\bot)$ and $\neg T(\bot)$, **PT**$^-$ without axioms for $\dot{\neg}$ derives the negation axioms for propositions and truth.

4.2 Generating **PT**-*Models*

Can we produce **PT**-models by generalized inductive definition? The difficulty is that the clause for introducing implication makes use (negatively) of the collection of truths. But we can adapt to our case a trick of Aczel [1].

Definition 26 Fix a model M of **TON**$^-$ and let X, Y range over subsets of the domain $|M|$ of M. X is called *suitable*, if $X_1 \subseteq X_0$ and for every u, if $u \in X$, then either $u = \langle 0, (u)_1 \rangle$ or $u = \langle 1, (u)_1 \rangle$.[10] We put $\mathcal{S} := \{X | X \text{ is suitable}\}$.

We use the following abbreviations:

1. $u \in X_0 := \langle 0, u \rangle \in X$ and $u \in X_1 := \langle 1, u \rangle \in X$.
2. A suitable X is determined by an ordered pairing $\langle X_0, X_1 \rangle$ of sets by letting $X_i := \{u | \langle i, u \rangle \in X\}$ where $i = 0, 1$.
3. We also define a partial ordering on suitable sets:

$$X \leq Y := X_0 \subseteq Y_0 \wedge \forall u (u \in X_0 \rightarrow (u \in X_1 \leftrightarrow u \in Y_1)). \quad (62)$$

4. If $\Gamma : \mathcal{S} \longrightarrow \mathcal{S}$, we define: $\Gamma_i(X) := \{u | \langle i, u \rangle \in \Gamma(X)\}$; Γ is $\leq$-monotone iff $X \leq Y$ implies $\Gamma(X) \leq \Gamma(Y)$.
5. An operator Γ is *suitable* if $\Gamma : \mathcal{S} \longrightarrow \mathcal{S}$ and Γ is $\leq$-monotone.

Hence by standard facts (see e.g. [7]) we can state the

[10] Recall footnote 2 of Sect. 2.4: think of $\langle a, b \rangle$, $(u)_0$, $(u)_1$ as values of the terms PAIRab, LEFTu, RIGHTu.

Lemma 27

(i) The structure $\langle \mathcal{S}, \leq \rangle$ is a partial ordering in which every $\leq$-increasing sequence of elements of $\mathcal{S}$ has a $\leq$-least upper bound in $\mathcal{S}$.
(ii) Every suitable operator Γ has fixed points (in particular there exists the $\leq$-least suitable X with $X = F(X)$).

We now define a suitable operator whose fixed points provide **PT**-models. This operator will be given by two separate operators described by means of elementary formulas in the language of **TON**, possibly expanded by parameters for naming subsets of M.

(i) $\mathcal{S}_0(u, X)$ is the formula

$$\begin{aligned} &\exists x \exists y((u = [x = y] \vee u = [N(x)]) \vee \\ &\vee (u = x \dot{\to} y \wedge X_0(x) \wedge (X_1(x) \to X_0(y))) \vee \\ &\vee (u = x \dot{\wedge} y \wedge X_0(x) \wedge X_0(y)) \vee \\ &\vee (u = [P(x)] \wedge X_0(x)) \vee \\ &\vee (u = \dot{\forall} x \wedge \forall z X_0(xz)))). \end{aligned}$$

(ii) $\mathcal{S}_1(u, X)$ is the formula

$$\begin{aligned} &\exists x \exists y((u = [x = y] \wedge x = y) \vee \\ &\vee (u = [N(x)] \wedge N(x)) \vee \\ &\vee (u = [P(x)] \wedge X_0(x)) \vee \\ &\vee (u = x \dot{\to} y \wedge X_0(u) \wedge (X_1(x) \to X_1(y))) \vee \\ &\vee (u = (x \dot{\wedge} y) \wedge X_0(u) \wedge X_1(x) \wedge X_1(y)) \vee \\ &\vee (u = (\dot{\forall} x) \wedge X_0(u) \wedge \forall u X_1(xu))). \end{aligned}$$

(iii) Finally, $\mathcal{S}(u, X)$—the disjoint sum of the two operators—is an elementary operator in the applicative language of **TON** with a new predicate variable X, which formalizes the clauses inductively generating the interpretation of T and P. Explicitly:

$$\mathcal{S}(t, X) \Leftrightarrow \exists i \exists y[t = (i, y) \wedge ((i = 0 \wedge \mathcal{S}_0(y, X)) \vee (i = 1 \wedge \mathcal{S}_1(y, X))].$$

$\mathcal{S}(u, B)$ is the formula, which result from $\mathcal{S}(u, Y))$ by replacing each subformula of the form $Y(t)$ with $B[x := t]$).

NB. Just to avoid further notational overloading, we identify operators with their defining formulas, and we leave the dependence on a fixed ground structure M implicit.

Lemma 28 *The operator defined by $\mathcal{S}$ is suitable (with respect to any M such that $M \models$ **TON**$^-$).*

Proof Clearly by definition the image of a suitable subset under $\mathcal{S}$ is suitable. Let us check that it preserves the ordering $\leq$. We argue informally. Let X, Y be suitable and $X \leq Y$. Then $X_0 \subseteq Y_0$ and hence, since $\mathcal{S}_0(a, X)$ is positive in X_0, $\mathcal{S}_0(a, Y)$. Thus it is enough to check

$$\mathcal{S}_0(a, X) \to (\mathcal{S}_1(a, X) \leftrightarrow \mathcal{S}_1(a, Y)). \tag{63}$$

There are several cases according to the form of a and all can be dealt with by standard arguments. Let us consider three cases.

1. Let $a = [P(x)]$, for some x. Then if we assume $\mathcal{S}_0(a, X)$, by definition of $\mathcal{S}_0$ we have $X_0(x)$, which trivially implies (63), since $\mathcal{S}_1(a, X) \equiv X_0(x)$, $\mathcal{S}_1(a, Y) \equiv Y_0(x)$ and $X_0 \subseteq Y_0$.
2. $a = x \dot{\to} y$. Assume $\mathcal{S}_0(a, X)$. Then $X_0(x)$ and $X_1(x) \to X_0(y)$. We want:
   - $\mathcal{S}_1(x \dot{\to} y, X) \to \mathcal{S}_1(x \dot{\to} y, Y)$;
   - $\mathcal{S}_1(x \dot{\to} y, Y) \to \mathcal{S}_1(x \dot{\to} y, X)$.

   As to the first implication, from the antecedent it follows $Y_0(x)$ since $X_0 \subseteq Y_0$. On the other hand if $X_1(x)$, X_1 and Y_1 coincide for elements of X_0; hence we conclude $Y_1(x)$, that is, we have shown $\mathcal{S}_1(x \dot{\to} y, Y)$. The second implication is similar.
3. Let $a = \dot{\forall} x$ and assume $\mathcal{S}_0(\dot{\forall} x, X)$, which implies $\forall u X_0(xu)$. By definition of the operator $\mathcal{S}_1$, we have to check $\forall u(X_1(xu)) \leftrightarrow \forall u(Y_1(xu))$. By logic, it is enough to prove $\forall u(X_1(xu) \leftrightarrow Y_1(xu))$. But if we choose any u with $X_0(xu)$, since $X \leq Y$, $X_1(xu) \leftrightarrow Y_1(xu)$. □

Theorem 29 *Let M be a model of* **TON**$^-$. *If X is a fixed point of the operator $\mathcal{S}$, then $\langle M, X\rangle \models$* **PT**$^-$. *If M is N-standard, $\langle M, X\rangle \models$* **PT**.

Proof Assume X satisfies

- $X_0(x) \leftrightarrow \mathcal{S}_0(x, X)$;
- $X_1(x) \leftrightarrow \mathcal{S}_1(x, X)$.

We have to show that every **PT**-axiom is satisfied, whenever we interpret $P(a)$, $T(a)$ by $X_0(a)$, $X_1(a)$ (in the given order).

Let us check the interpretation of $T(a) \to P(a)$. So assume $X_1(a)$; since X is a fixed point, $\mathcal{S}_1(a, X)$. There are several cases according to the form of a. If $a = [x = y]$ or $a = [N(x)]$, by definition of $\mathcal{S}_0$, we have $\mathcal{S}_0(a, X)$, and hence $X_0(a)$. Let $a = [P(x)]$; then $X_0(x)$, i.e. again by definition of $\mathcal{S}_0$, $\mathcal{S}_0([P(x)], X)$ whence, since X is a fixed point, $X_0([P(x)])$. The converse is similar. In all other cases, by inspection of $\mathcal{S}_1(a, X)$, $X_0(a)$ follows.

Consider the interpretation of $T(P(a)) \leftrightarrow P(a) \leftrightarrow P([P(x)])$. Indeed, by definition and fixed point property:

$$\begin{aligned} X_1([P(x)]) &\Leftrightarrow \mathcal{S}_1([P(x)], X) \\ &\Leftrightarrow X_0(x) \\ &\Leftrightarrow \mathcal{S}_0([P(x)], X) \\ &\Leftrightarrow X_0([P(x)]). \end{aligned} \tag{64}$$

Consider the interpretation of $P(x \dot{\rightarrow} y) \leftrightarrow P(x) \wedge (T(x) \rightarrow P(y))$. Indeed by fixed point and definition of $\mathcal{S}_0$, $\mathcal{S}_1$:

$$\begin{aligned} X_0(x \dot{\rightarrow} y) &\Leftrightarrow \mathcal{S}_0(x \dot{\rightarrow} y, X) \\ &\Leftrightarrow X_0(x) \wedge (X_1(x) \rightarrow X_0(y)). \end{aligned} \tag{65}$$

Let us check the soundness of $\forall u T(xu) \rightarrow T(\dot{\forall} x)$. Then:

$$\begin{aligned} \forall u X_1(xu) &\Rightarrow \forall u \mathcal{S}_1(xu, X) \\ &\Rightarrow \mathcal{S}_1(\dot{\forall} x, X) \\ &\Rightarrow X_1(\dot{\forall} x). \end{aligned} \tag{66}$$

The remaining cases are also straightforward. □

4.3 Upper Bounds for **PT**⌈ *and* **PT**

In order to classify the proof-theoretic strength of **PT**, **PT**⌈, we consider variants, which serve for proof theoretic investigations. We formalize the new systems in the sublanguage $\mathcal{L}_t$ of $\mathcal{L}_T$ without the predicate P, but we adopt the obvious abbreviation

$$P(x) := T(x) \vee T(\dot{\neg} x) \tag{67}$$

so that we can identify the new language with the language of **PT**.

The basic *positive atoms* have the form: $t = s, N(t), T(t)$. The *negative atoms* are obtained by negating the positive ones; an *atom* is simply a positive or a negative atom and we stipulate that $\neg\neg A := A$ (A atom). Formulas are inductively generated from atoms by closing under disjunction, conjunction, unbounded quantification. If A is an arbitrary formula, $\neg A$ is the formula which results from the negation normal form of $\neg A$ by erasing each even sequence of occurrences of negation in front of atoms.

If $Q := P, T$, a formula A of $\mathcal{L}_T$ is *Q-positive* (*Q-negative*) if every occurrence of Q in A occurs within positive (negative) atoms of the form $Q(t)$ ($\neg Q(t)$). A formula

A is *Q-separated* if A is Q-positive or Q-negative. A formula A is *Q-free* if Q does not occur in A. A Q-free formula can be regarded as both Q-positive and Q-negative.

The *rank* of a formula over its Q-separated formulas is assigned as follows: a) if A is Q-separated, $rk(A) = 0$; b) else, if A or B is not Q-separated, $rk(A \circ B) = \max(rk(A), rk(B)) + 1$ ($\circ$ is a conjunction or a disjunction); $rk(QsA) = rk(A) + 1$ (where Qs is an unbounded quantifier).

Definition 30

(i) $\mathcal{T}_w(u, Y)$ is an elementary positive operator[11] in the applicative language of **TON**, expanded with a new predicate variable Y; $\mathcal{T}_w(u, Y)$ formalizes the clauses inductively generating the interpretation of T, and Y occurs positively in it. Explicitly:

$$
\begin{aligned}
\mathcal{T}_w(t, Y) \quad \Leftrightarrow \quad & \exists x \exists y((t = [x = y] \wedge x = y) \vee \\
& \vee (t = [\neg x = y] \wedge \neg x = y) \vee \\
& \vee (t = [N(x)] \wedge N(x)) \vee \\
& \vee (t = [\neg N(x)] \wedge \neg N(x)) \vee \\
& \vee (t = \dot{\neg}(\dot{\neg} x) \wedge Y(x)) \vee \\
& \vee (t = (x \dot{\wedge} y) \wedge Y(x) \wedge Y(y)) \vee \\
& \vee (t = \dot{\neg}(x \dot{\wedge} y) \wedge ((Y(\dot{\neg} x) \wedge Y(y)) \vee \\
& \vee (Y(\dot{\neg} y) \wedge Y(x)) \vee (Y(\dot{\neg} x) \wedge Y(\dot{\neg} y))) \vee \\
& \vee (t = x \dot{\rightarrow} y \wedge ((Y(x) \wedge Y(y)) \vee Y(\dot{\neg} x))) \vee \\
& \vee (t = \dot{\neg}(x \dot{\rightarrow} y) \wedge ((Y(x) \wedge Y(\dot{\neg} y)))) \vee \\
& \vee (t = (\dot{\forall} x) \wedge \forall u Y(xu)) \vee \\
& \vee (t = \dot{\neg}(\dot{\forall} x) \wedge \exists u Y(\dot{\neg}(xu)) \wedge \forall u (Y(xu) \vee Y(\dot{\neg}(xu))))).
\end{aligned}
$$

$\mathcal{T}_w(u, B)$ is the formula, which result from $\mathcal{T}_w(u, Y))$ by replacing each subformula of the form $Y(t)$ with $B[x := t]$).

Definition 31 **FL**$\lceil$[12] is consists of

1. logical axioms of the form

$$
\begin{aligned}
& \Gamma, \neg A, A \\
& \Gamma, \neg t = s, A[x := t], A[x := s]
\end{aligned}
$$

   where A is an atom (according to the previous definitions);
2. axioms of the form Γ, Δ where Δ is an e-atom or a finite set of e-atoms; Δ formalizes the standard axioms for extended combinatory, logic, natural numbers;

[11] In the standard sense, see [25].

[12] **FL** is reminiscent of Feferman's logic.

3. standard logical rules for introducing $\wedge$, $\vee$, $\forall$, $\exists$ and the cut rule

$$\frac{\Gamma, A \quad \Gamma, \neg A}{\Gamma}$$

4. T-consistency:

$$\Gamma, \neg T(t), \neg T(\dot{\neg} t)$$

5. N-induction rule for *determinate functions* (see Sect. 2.2):

$$\frac{\Gamma, PF(f) \quad \Gamma, T(f0) \quad \Gamma, \forall x(N(x) \rightarrow (T(fx) \rightarrow T(f(x+1))))}{\Gamma, \neg N(t), T(ft)}$$

6. T-closure:

$$\frac{\Gamma, \mathcal{T}_w(t, T)}{\Gamma, T(t)}$$

7. T-soundness:

$$\frac{\Gamma, \neg \mathcal{T}_w(t, T)}{\Gamma, \neg T(t)}$$

FL is the calculus which is obtained from $\mathbf{FL}\lceil$ by replacing N-induction rule for *determinate functions* by full N-induction rule, that is, if $A(x)$ is an arbitary formula of $\mathcal{L}_t$,

$$\frac{\Gamma, A(0) \quad \Gamma, \forall x(N(x) \rightarrow (A(x) \rightarrow A(x+1)))}{\Gamma, \neg N(t), A(t)}$$

Remark 7 The label **FL** hints at a compositional theory of truth introduced by Kentaro Fujimoto in [13] under the label **FKF**. The original theory **FKF** is a variant of Kripke-Feferman **KF** over Peano arithmetic, where one assumes for the predicate T Feferman's logic for determinate truth [10].

Definition 32 If $S :=$ $\mathbf{FL}\lceil$, **FL**, **TON**, we inductively define a derivability relation $S \vdash^m_n \Gamma$ as follows:

(i) If Γ is an axiom, $S \vdash^m_n \Gamma$;
(ii) assume Γ is inferred by means of a finitary rule (but not a cut) from the set $\{\Gamma_i | i \leq j\}$; if $S \vdash^{m_i}_{n_i}$ where $m_i < m$, $n_i \leq n$, $i \leq j$, then also $S \vdash^m_n \Gamma$;
(iii) assume Γ is inferred by means of a cut from the premises Γ, A and $\Gamma, \neg A$. If $S \vdash^w_u \Gamma, A$ and $S \vdash^p_q \Gamma, \neg A$ where $w, p < m$ and $R(A) + 1, u, q \leq n$, $S \vdash^m_n \Gamma$.

Lemma 33 (Partial Cut-elimination) *Let* $\mathbf{FL}\lceil \vdash^m_{k+2} \Gamma$. *Then* $\mathbf{FL}\lceil \vdash^{2^m}_{k+1} \Gamma$. *Hence every* $\mathbf{FL}\lceil$*-derivation* $\mathcal{D}$ *can be effectively transformed into a* $\mathbf{FL}\lceil$*-derivation of the same end-sequent, where cut-formulas are* T*-separated and have rank* 0.

Proof The argument is standard; it essentially depends on the fact that the active formulas in the axioms and in the conclusions of the mathematical inferences are T-separated. □

Lemma 34 *The system* $\mathbf{FL}^-$ *proves*

$$\begin{array}{rcl} T(x) & \to & P(x), \\ P([x=y]) & \wedge & P([N(x)]), \\ P(x) & \to & \neg(T(x) \wedge T(\dot{\neg} x)), \\ T(\dot{\neg}\dot{\neg} x) & \leftrightarrow & T(x), \\ T(x \dot{\wedge} y) & \leftrightarrow & T(x) \wedge T(y), \\ T(\dot{\neg}(x \dot{\wedge} y)) & \leftrightarrow & (T(\dot{\neg} x) \wedge T(\dot{\neg} y)) \vee, \\ & & \vee\, (T(\dot{\neg} x) \wedge T(y)) \vee (T(x) \wedge T(\dot{\neg} y)), \\ T(x \dot{\to} y) & \leftrightarrow & (T(x) \wedge T(y)) \vee T(\dot{\neg} x), \\ T(\dot{\neg}(x \dot{\to} y)) & \leftrightarrow & (T(x) \wedge T(\dot{\neg} y)), \\ T(\dot{\forall} f) & \leftrightarrow & \forall x T(fx), \\ T(\dot{\neg}\dot{\forall} f) & \leftrightarrow & \exists x T(\dot{\neg}(fx)) \wedge \forall x P(fx). \end{array} \tag{68}$$

Proof Apply T-consistency, and T-closure, T-soundness rules. □

Lemma 35 $\mathbf{PT}^- \vdash A$ *iff* $\mathbf{FL}^- \vdash A$.

Proof This amounts to check that the axioms ruling implication internally are interdeducible. But this is an exercise in formal deducibility. □

4.3.1 Approximating Truth by Its Finite Levels

Is it possible to approximate truth by its finite levels and hence to eliminate truth?

Definition 36 Let $\bot = \mathsf{K} = \mathsf{S}$; then

$$T^0(t) = \bot, \tag{69}$$

$$T^{m+1}(t) = \mathcal{T}_w(t, T^m)). \tag{70}$$

Clearly each formula in the sequence belongs to the language $\mathcal{L}_{op}$.

If A is any formula in negation normal form, let $A[m, n]$ be obtained from A by replacing each atom of the form $T(t)$ $(\neg T(t))$ by $T^n(t)$ $(\neg T^m(t))$. Clearly $T^m(t)$, $\neg T^n(t)$ are T-free formulas of $\mathcal{L}_{op}$.

Lemma 37 *If* $0 < m_2 \leq m_1 \leq k_1 \leq k_2$, Γ *is a set of formulas such that* $\mathbf{TON} \vdash^n_p \Gamma[m_1, k_1], \Delta$, *then* $\mathbf{TON} \vdash^n_p \Gamma[m_2, k_2], \Delta$.

Theorem 38 *Let* $\mathcal{D}$ *be a* $\mathbf{FL}\lceil$*-derivation of* Γ *with height k, where cut-formulas are T-separated. Then, provably in* $\mathbf{TON}$*, for every* $m > 0$,[13] *if* $H(m) := m + 2^k$*:*

$$\Gamma[m, H(m)]. \tag{71}$$

[13] In general, if $\Gamma := \{A_1, \ldots, A_q\}$, $\Gamma[m, n] := \{A_1[m, n], \ldots, A_q[m, n]\}$.

Proof The restriction to propositional functions has the effect that N-induction for $\mathcal{L}_{op}$-formulas is enough. The asymmetric interpretation relies on the fact that cut formulas are always T-separated. Of course, note that the transform $A \mapsto A[m, n]$ is the identity function on formulas of $\mathcal{L}_{op}$.

1. Logical axioms and rules: by persistence.
2. Cut: use partial cut elimination and a standard substitution argument.
3. Consistency: verify by outer induction on m

$$\forall x(\neg T^m(x) \vee \neg T^m(\dot{\neg}x)). \tag{72}$$

4. N-induction rule for *determinate functions*: assume that, provably in **TON**, for some g, given any $m > 0$, we have $m \leq g(m)$ and

$$T^{g(m)}(f0), \tag{73}$$
$$\forall x(N(x) \rightarrow (T^m(fx) \rightarrow T^{g(m)}(f(x+1))), \tag{74}$$
$$\forall x(T^{g(m)}(fx) \vee T^{g(m)}(\dot{\neg}(fx))). \tag{75}$$

   Then by $\mathcal{L}_{op}$-IND_N it is enough to verify:

$$\forall x(N(x) \rightarrow (T^{g(m)}(fx) \rightarrow T^{g(m)}(f(x+1))).$$

   Assume $N(x)$, $T^{g(m)}(fx)$: then by (74),

$$T^{g(g(m))}(f(x+1)). \tag{76}$$

   Moreover by (75), consistency and downward persistence, we have

$$T^{g(m)}(f(x+1)) \vee T^{g(m)}(\dot{\neg}(f(x+1)), \tag{77}$$
$$\neg T^{g(g(m))}(f(x+1)) \vee \neg T^{g(m)}(\dot{\neg}(f(x+1))). \tag{78}$$

   whence by cut between (77) and (78)

$$T^{g(m)}(f(x+1)) \vee \neg T^{g(g(m))}(f(x+1)). \tag{79}$$

   The conclusion follows again by cut between (79) and (76).
5. T-closure (T-soundness): straightforward. □

Corollary 39 *If* **PT**$\lceil\vdash A$ *and* $A \in \mathcal{L}_{op}$, *then* **TON** $\vdash A$.

Proof Apply Theorem 38. □

Let $\Pi^0_1 - CA_{<\varepsilon_0}$ be the theory of iterated jump up to any $\alpha < \varepsilon_0$.[14]

[14]For a precise definition, see [11].

Theorem 40

(i) $\mathbf{PA} \equiv \mathbf{PT}\lceil \equiv \mathbf{CT}\lceil$.
(ii) $\Pi^0_1 - CA_{<\varepsilon_0} \equiv \mathbf{PT}$.

Proof

(i) Ad $\mathbf{PA} \leq \mathbf{PT}\lceil$: obvious.
(ii) Ad $\mathbf{PT}\lceil \leq \mathbf{PA}$: by Theorem 38 since $\mathbf{CT}\lceil$ is a subtheory of $\mathbf{PT}\lceil$(even if $\dot{\neg}$ is omitted, see by Remark 6).
(iii) Ad $\Pi^0_1 - CA_{<\varepsilon_0} \leq \mathbf{PT}$: consider the ω-model consisting—as range of second order variables—of the collection PF_N of propositional functions which define subsets of N. Then the jump hierarchy up to any $\alpha < \varepsilon_0$ can be shown to exist in PF_N by applying Lemma 25.
(iv) Ad $\mathbf{PT} \leq \Pi^0_1 - CA_{<\varepsilon_0}$: by lifting the previous method to the case of systems with full number theoretic induction. In this case we rely upon a standard embedding into systems with ω-rule. □

Remark 8 We can strengthen the theorem above by adding generalized induction schemata over truth, e.g. in the form of the following *inference rule of T-induction*: if $B(x)$ is any *T-positive* $\mathcal{L}_T$-formula,

$$\frac{\Gamma, \forall x(\mathcal{T}_w(x, B) \to B(x))}{\Gamma, \neg T(t), B(t)} \tag{80}$$

Then **GID**-induction can be eliminated in favour of a suitable infinitary rule $\mathbf{T}^\infty$:

$$\frac{\Gamma, \neg T^n(t) \quad \text{for each } n \in \omega}{\Gamma, \neg T(t)} \tag{81}$$

The rule $\mathbf{T}^\infty$ allows to show each instance of the schema of generalized on truth (argue by induction on $n \in \omega$). Moreover the system based on (81)—with N-induction restricted to determinate functions—satisfies partial cut elimination theorem. Of course, derivation trees are now infinitary and the interpretation theorem holds with $m \mapsto H(m)$, where H is an α-recursive function,[15] for some $\alpha < \varepsilon_0$.

Remark 9 Feferman [10] stated a conjecture about the strength of the theory of determinate truth over **PA**, and the conjecture has been settled by Fujimoto [13] by employing the so-called relative truth definability. The construction above can be adapted to the system **DT** over **PA**, in order to produce an alternative proof of the conjecture.

[15]See [26, 27].

Remark 10 In view of Theorems 6 and 40, **PT** is strictly stronger than **CT** and hence the proof theoretic strength of **PT** over **CT** resides in the implication axioms (55) and (56). This should be contrasted with the situation in [13], Theorem 50.

4.4 Adding the Join Operator

4.4.1 Axioms for Explicit Types: Join

Define:

- $\mathcal{R}(a) := \exists Y \mathcal{R}(a, Y)$.
- $s \dot{\in} t := \exists Y(\mathcal{R}(t, Y) \wedge s \in Y)$.
- Join is the principle $(\mathcal{J})$:

$$\mathcal{R}(a) \wedge \forall x(x \dot{\in} a \rightarrow \mathcal{R}(fx)) \rightarrow \mathcal{R}(\mathsf{j}\,(a,f)) \wedge \wedge \forall u, v(\langle u, v\rangle \dot{\in} \mathsf{j}\,(a,f) \leftrightarrow u \dot{\in} a \wedge v \dot{\in} fu)), \tag{82}$$

- **EETJ**:= **EET** with $(\mathcal{J})$.
- **EETJ**$\lceil$: as **EETJ** except that it now includes only type induction for numbers (24).

Theorem 41 **EETJ**(**EETJ**$\lceil$) *is interpretable in* **PT** (**PT**$\lceil$).

Proof Apply Lemma 25 and the previous interpretability results about **EET**, since **PT**$^-$ contain **CT**$^-$. ☐

5 Truth and Types IV

5.1 Abstract 'Kripke-Feferman'

We outline an abstract version of the Kripke-Feferman system over **PA**, which can also be regarded as the theory of *classical* Frege structures (see [6, 14, 22]).

Definition 42 **KF**$\lceil$ comprises the base theory **TON**$^-$, and the fixed point axiom $(\mathcal{T})$ for abstract truth:

$$\mathcal{T}(x, T) \leftrightarrow T(x). \tag{83}$$

Here $\mathcal{T}(x, T)$ is a formula encoding the closure properties:

$$\frac{a = b}{T[a = b]} \qquad \frac{\neg(a = b)}{T[\neg(a = b)]} \qquad \frac{N(a)}{T[N(a)]} \qquad \frac{\neg N(a)}{T[\neg N(a)]}$$

for the basic atomic formulas with = and N. Further, the following additional clauses for the compound formulas:

$$\frac{T(a)}{T(\dot{\neg}\dot{\neg}a)} \qquad \frac{T(a) \quad T(b)}{T(a \dot{\wedge} b)} \qquad \frac{T(\dot{\neg}a)\ [\ or\ T(\dot{\neg}b)]}{T(\dot{\neg}(a \dot{\wedge} b))}$$

$$\frac{\forall x\, T(ax)}{T(\dot{\forall}a)} \qquad \frac{\exists x\, T(\dot{\neg}ax)}{T(\dot{\neg}\dot{\forall}a)}$$

Finally $\mathbf{KF}\lceil$ includes:

1. Consistency axiom: $\neg(T(x) \wedge T(\dot{\neg}x))$.
2. the axiom $\mathsf{D\text{-}IND}_N$ of induction on natural numbers N for functions with determinate truth values (see Sect. 2.3).

5.1.1 Recursion-Theoretic Structure

For each formula A, if $\mathbf{Y}$ is the Curry fixed point combinator, define

$$\mathbf{I}(A) \ := \ \mathbf{Y}(\lambda v.\{x : A(x, v)\}).$$

Hence by β-conversion $\mathbf{I}(A) = \{x : A(x, \mathbf{I}(A))\}$, and a general *second recursion theorem* for predicates holds in $\mathbf{KF}^{-}$:

Lemma 43 *If A is T-positive*

$$\forall x\,(T(\mathbf{I}(A)x) \ \leftrightarrow \ A(x, \mathbf{I}(A))). \tag{84}$$

Theorem 44 *Let $\mathcal{M}$ be a model of* **TON** *and let MIN_M be the least fixed point model of $\mathbf{KF}_\mu\lceil$ expanding $\mathcal{M}$. Then $\mathbf{I}(A)$ represents the least fixed-point of the monotone operator defined by A in MIN_M.*[16]

This suggests the schema **GID**, ensuring the minimality of the fixed points: if $A(x, v)$ is a positive operator

$$\mathbf{Clos}_A(B) \rightarrow \forall x\,(T(\mathbf{I}(A)x) \rightarrow B(x))$$

with $\mathbf{Clos}_A(B) \ := \ \forall x\,(A(x, B) \rightarrow B(x))$.

Theorem 45 ([6])

(i) $\mathbf{KF}\lceil+\mathbf{GID}$ *is proof-theoretically equivalent to* **PA**.
(ii) $\mathbf{KF} + \mathbf{GID}$ *is proof-theoretically equivalent to* $\mathbf{ID_1}$.

[16] This means: the set of all $a \in \mathcal{M}$ satisfying $T(I(A)x$ in MIN_M is the least fixed point of the operator defined by A in MIN_M.

5.2 Partial Truth with Minimality

$\mathbf{KF}_\mu$ (Truth with minimality, see Burgess [2]): it is the fragment of $\mathbf{KF}$[17] with

(i) only the *composition principles*, e.g. $\forall x(\mathcal{T}(x, T) \to T(x))$;
(ii) the schema: if B is an arbitrary formula,

$$\forall x(\mathcal{T}(x, B) \to B(x)) \to \forall x(T(x) \to B(x)). \tag{85}$$

Then $\mathbf{KF}_\mu$ proves the decomposition axioms $\forall x(T(x)) \to \mathcal{T}(x, T))$ and the consistency axiom. Also, it explicitly refutes statements that fail in the least fixed point model, e.g. the so-called Truth-Teller S such that, provably in $\mathbf{KF}^-$, $S = \dot{T}S$. Indeed, choose $T_S(x) := T(x) \wedge x \neq S$. Then it's easy to check by independence and T-closure that $\forall x(\mathcal{T}(x, T_S) \to T_S(x)$, and hence $T(x)$ implies $T_S(x)$, for arbitrary x. Therefore $T(S)$ implies $T_S(S)$, whence $\neg T(S)$ by logic.

Clearly $\mathbf{KF}_\mu\lceil$ has an inner model in $\mathbf{KF}\lceil+\mathbf{GID}$.

As to the upper bound, we can adapt the direct proof theoretic analysis of $\mathbf{KF}\lceil+\mathbf{GID}$ (see [3]) to $\mathbf{KF}_\mu\lceil$.

As to the lower bound, let $\mathbf{ID}_\mathbf{1}^\mathbf{acc}$ be the theory of accessibility inductive definitions over $\mathbf{PA}$, which is known to be proof-theoretically equivalent to the theory of elementary inductive definitions. Let $\prec$ be a binary relation encoded by a propositional function,[18] and let $Field(\prec) := \{u | \exists v(\langle u, v\rangle \eta \prec \vee \langle v, u\rangle \eta \prec)\}$. Let

$$\Phi := \lambda a \lambda x.[\forall y(y \prec x \to ay)]$$

where $[\forall y(y \prec u \to ay)]$ is the applicative term $\dot{\forall}(\lambda y.(\dot{\to})[y \prec x](ay))$. Then we can choose by (84) a term $W(\prec)$, such that—using the notations of Sect. 2.2—$\mathbf{KF}\lceil$ proves

$$\begin{aligned} W(\prec)x &= [\forall y(y \prec x \to W(\prec)y)], \\ x\eta W(\prec) &\leftrightarrow x\eta Field(\prec) \wedge \forall y(y \prec x \to y\eta W(\prec)). \end{aligned}$$

The schema of transfinite induction along $\prec$ has the form

$$TI(\prec, A) := Progr(\prec, A) \to \forall x(x\eta W(\prec) \to A(x)). \tag{86}$$

where

$$Progr(\prec, A) := (\forall x\eta Field(\prec))(\forall y(y \prec x \to A(y)) \to A(x)).$$

---

[17] As for $\mathbf{KF}$, a warning: we keep using the same label of [2] for a theory $\mathbf{KF}_\mu$, which is *not* an extension of Peano Arithmetic.

[18] So $\prec$ is determinate in the sense of (2).

Then we lift to the present context the (classical form of so-called) *bar-induction* schema, which goes back to Kreisel; the proof below takes inspiration from an analogous result of [15], also exploited by [2]:

Theorem 46 *If A is an arbitrary formula, $\prec$ is a (propositional function encoding a) binary relation, then the schema of transfinite induction on the largest well-founded part $W(\prec)$ of $\prec$ holds, provably in $\mathbf{KF}_\mu$.*

Proof For simplicity, it is convenient to work in a definitional extension of $\mathbf{KF}_\mu$, where $T(\dot{\neg}x) \vee T(y) \to T(x \dot{\to} y)$ is provable. Fix any A, $\prec$, such that $\prec$ is a propositional function. Assume that A is $\prec$-progressive and $u\eta W(\prec)$. It is sufficient to verify $A(u)$. For the sake of semplification, we further assume that the field of $\prec$ is the whole universe (so that we can avoid to make explicit reference to it). Define

$$\begin{aligned} T_A(x) :\Leftrightarrow\ & T(x) \wedge \forall u(x = W(\prec)u \to A(u)) \wedge \\ & \wedge \forall u \forall v(x = [v \prec u \to W(\prec)v] \wedge v \prec u \to A(v)). \end{aligned} \tag{87}$$

Assume that we have shown

$$\forall x(\mathcal{T}(x, T_A) \to T_A(x)). \tag{88}$$

Then by the schema (85) we can conclude $\forall x(T(x) \to T_A(x))$. Hence by assumption, for $x := W(\prec)u$, we have $T_A(W(\prec)u)$, which immediately implies $A(u)$.

Hence it remains to check (88). But this follows with the closure axioms of T and with the independence conditions ruling the dotted constants $\dot{N}$, $\dot{=}$, etc. Let us consider three cases.

1. Assume $\mathcal{T}([a = b], T_A)$; then $a = b$ holds and hence $T([a = b])$. Since $W(\prec)u = \dot{\forall} f$ for a suitable f, then $W(\prec)u \neq [a = b]$ and

$$[v \prec u \to W(\prec)v] \neq [a = b]$$

   by (3). So we can trivially conclude $T_A([a = b])$.
2. Assume $\mathcal{T}(\dot{\forall} f, T_A)$; then for all x, $T_A(fx)$. Then by definition of T_A, we have $\forall x T(fx)$ whence $T(\dot{\forall} f)$. It remains to check:

$$\forall u(\dot{\forall} f = W(\prec)u \to A(u)), \tag{89}$$

$$\forall u \forall v(\dot{\forall} f = [v \prec u \to W(\prec)v] \wedge v \prec u \to A(v)). \tag{90}$$

   The second is trivially true by independence. As to the first condition, assuming $\dot{\forall} f = W(\prec)u$, we must prove $A(u)$. But $\dot{\forall} f = W(\prec)u$ implies

$$f = \lambda v[v \prec u \to W(\prec)v],$$

whence, for arbitrary v:

$$fv = [v \prec u \to W(\prec)v]. \tag{91}$$

Since for all x, $T_A(fx)$, we have, for $x := v$

$$\forall y \forall z(fv = [y \prec z \to W(\prec)y] \wedge y \prec z \to A(y)).$$

Hence, for $y := v$, $z := u$

$$fv = [v \prec u \to W(\prec)v] \wedge v \prec u \to A(v).$$

By (91), for arbitrary v :

$$v \prec u \to A(v).$$

But A is $\prec$-progressive and hence $A(u)$. It follows $T_A(\dot{\forall} f)$.

3. Assume $T_A(\dot{\neg} a) \vee T_A(b)$. We check $T_A(a \dot{\to} b)$. By assumption $T(\dot{\neg} a) \vee T(b)$ and hence $T(a \dot{\to} b)$. So it is enough to check

$$\forall u(a \dot{\to} b = W(\prec)u \to A(u)), \tag{92}$$

$$\forall u \forall v(a \dot{\to} b = [v \prec u \to W(\prec)v] \wedge v \prec u \to A(v)). \tag{93}$$

(92) is trivial by independence (since $\dot{\to} \neq \dot{\forall}$). As to (93), assume $v \prec u$ and $a \dot{\to} b = [v \prec u \to W(\prec)v]$. Then $a = [v \prec u]$ and $b = W(\prec)v$. Were $T_A(\dot{\neg} a)$, then $T([\neg v \prec u])$; but $\prec$ is a propositional function and hence $\neg v \prec u$, contradiction! Hence $T_A(b)$, and, in particular:

$$\forall u \forall v(b = W(\prec)v \wedge v \prec u \to A(v)).$$

By separation $A(v)$, and we have checked (93). □

Hence by the previous lemma:

Corollary 47 $\mathbf{ID}_1^{\mathbf{acc}}$ *is interpretable in* $\mathbf{KF}_\mu$.

Conjecture 2 $\mathbf{KF}\lceil + \mathbf{GID}$ *is interpretable in* $\mathbf{KF}_\mu\lceil$.

Note that, according to [2], the conjecture holds for the ordinary formal system formalized in the language of Peano arithmetic.

5.3 Explicit Types and Name Induction

There is a simple extension of **EETJ** which corresponds to truth minimality in $\mathbf{KF}_\mu$. The idea (see [23]) is that *names are inductively generated from the basic constructors only* (identity id, natural numbers nat, inverse image inv, domain dom, complement co, intersection int, join j). Therefore, if $B(x)$ satisfies the same closure conditions $\mathcal{N}(x, -)$[19] as the name constructors, B contains all names:

$$\forall x(\mathcal{N}(x, B) \rightarrow B(x)) \rightarrow \forall x(\mathcal{R}(x) \rightarrow B(x)). \tag{94}$$

NEM := **EETJ** + (94).

Theorem 48 $\mathbf{ID_1}$*,* **NEM***,* $\mathbf{KF}_\mu$*,* **KF** + **GID** *are proof-theoretically equivalent.*

As to the proof, the crucial step is Theorem 1 in [23], that is, if $\prec$ is a binary relation on X, the largest well-founded part $W(\prec, X)$ of $\prec$ can be assigned a name, uniformly in any given name for X and $\prec$. Hence $\mathbf{ID_1^{acc}}$ is interpretable in **NEM**. On the other hand, it is straighforward to check:

Lemma 49 **NEM** *is interpretable into* **KF** + **GID**.

References

1. P. Aczel, Frege structures and the notions of proposition, truth and set, in *The Kleene Symposium*, ed. by J. Barwise, H.J. Keisler, K. Kunen (North Holland, Amsterdam, 1980), pp. 31–59
2. J. Burgess, Friedman and the axiomatization of Kripke's theory of truth, in *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, ed by N. Tennant (College Publications, London, 2014), pp. 125–148
3. A. Cantini, Levels of implications and type free theories of partial classifications with approximation operator. Zeitschrift f. Math. Logik u. Grundlagen **38**, 107–141 (1992)
4. A. Cantini, P. Minari, Uniform inseparability in explicit mathematics. J. Symbolic Logic **61**, 313–326 (1999)
5. A. Cantini, On a Russellian paradox about propositions and truth, in *One Hundred Years of Russell's Paradox. Mathematics, Logic and Philosophy* ed by G. Link (Walter de Gruyter, Berlin, 2004), pp. 259–284
6. A. Cantini, *Logical Frameworks for Truth and Abstraction, Studies in Logic and the Foundations of Mathematics*, vol. 135 (North Holland, Amsterdam, 1996)
7. B.A. Davey, H.A. Priestley, *Introduction to Lattices and Order*, 2nd edn. (Cambridge, 2002)
8. S. Eberhard, T. Strahm, Weak theories of truth and explicit mathematics, in *Logic, Construction, Computation*, ed. by U. Berger, H. Diener, P. Schuster (Ontos Verlag, 2012), pp. 156–183
9. S. Feferman, Constructive theories of operations and classes, in *Logic Colloquium '78* ed by M. Boffa, et al. (North-Holland, 1979), pp. 159–224
10. S. Feferman, Axioms for determinate truth. Rev. Symbolic Logic **1**, 204–217 (2008)
11. S. Feferman, Reflecting on incompleteness. J. Symbolic Logic **56**, 1–49 (1991)
12. S. Feferman, Does reductive proof theory have a viable rationale? Erkenntnis **53**, 63–96 (2000)

[19] We will not spell them explicitly for the sake of brevity.

13. K. Fujimoto, Relative truth definability of axiomatic truth theories. Bull. Symbolic Logic **16**, 305–344 (2010)
14. R. Flagg, J. Myhill, Implication and analysis in classical Frege structures. Ann. Pure Appl. Logic **34**, 33–85 (1987)
15. H. Friedman, M. Sheard, An axiomatic approach to self-referential truth. Ann. Pure Appl. Logic **33**, 1–21 (1987)
16. T. Glass, On power set in explicit mathematics. J. Symbolic Logic **61**, 468–489 (1996)
17. V. Halbach, *Axiomatic Theories of Truth* (Cambridge University Press, Cambridge, 2011)
18. G. Jäger, Type theory and explicit mathematics, in *Logic Colloquium '87*, Studies in Logic and the Foundations of Mathematics, vol. 129 (North-Holland, Amsterdam, 1989), pp. 117–135
19. G. Jäger, Power types in explicit mathematics? J. Symbolic Logic **62**, 1141–1146 (1997)
20. G. Jäger, R. Kahle, A. Setzer, T. Strahm, The proof theoretic analysis of transfinitely iterated fixed point theories. J. Symbolic Logic **64**, 53–67 (1999)
21. G. Jäger, T. Strahm, Totality in applicative theories. Ann. Pure Appl. Logic **74**, 105–120 (1995)
22. R. Kahle, Truth in applicative theories. Studia Logica **68**, 103–128 (2001)
23. R. Kahle, T. Studer, A theory of explicit mathematics equivalent to ID_1, in *Computer Science Logic (Fischbachau)*, Lecture Notes in Computer Science, 1862 (Springer, Berlin, 2000), pp. 356–370
24. G.E. Leigh, Conservativity for theories of compositional truth via cut elimination. J. Symbolic Logic, **80**, 845–865 (2015)
25. Y.N. Moschovakis, *Elementary Induction on Abstract Structures, Studies in Logic and the Foundations of Mathematics*, vol. 77 (North Holland, Amsterdam, 1974)
26. W. Pohlers, *Proof Theory* (Springer, Berlin-New York, 2009)
27. H. Schwichtenberg, S. Wainer, in *Proofs and Computations*, Perspectives in Logic (ASL and Cambridge University Press, Cambridge, 2012)
28. T. Strahm, Theories with self-application and computational complexity. Inf. Comput. **185**, 263–297 (2003)

http://www.springer.com/978-3-319-29196-3

Advances in Proof Theory
Kahle, R.; Strahm, Th.; Studer, Th. (Eds.)
2016, XII, 425 p. 10 illus., Hardcover
ISBN: 978-3-319-29196-3
A product of Birkhäuser Basel