UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE
Repository istituzionale dell'Università degli Studi di Firenze

Reflecting and unfolding, in: Ademollo, F., Amerini, F. and De Risi, V. (editors). Thinking and Calculating: Essays in Logic, its History and

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Reflecting and unfolding, in: Ademollo, F., Amerini, F. and De Risi, V. (editors). Thinking and Calculating: Essays in Logic, its History and its Philosophical Applications in Honour of Massimo Mugnai, pp. 1-15, Cham: Springer International Publishing, ISBN:9783030973025, / Francesco Ademollo, Pierluigi Minari, Andrea Cantini. - STAMPA. - Logic, Epistemology and the Unity of Science 54:(2023), pp. 1-456.

Availability:

This version is available at: 2158/1350272.2 since: 2024-02-16T08:45:39Z

Publisher:

Ademollo, Francesco

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:

Conformità alle politiche dell'editore / Compliance to publisher's policies

Questa versione della pubblicazione è conforme a quanto richiesto dalle politiche dell'editore in materia di copyright.
This version of the publication conforms to the publisher's copyright policies.

(Article begins on next page)

02 May 2024

Reflecting and unfolding

Andrea Cantini

Notionis distinctae primitivae non alia datur cognitio, quam intuitiva, ut compositarum plerumque cogitatio non nisi symbolica est (G.W. Leibniz, 1684[1])

Abstract

The philosophical problem of *implicit commitment* can be roughly stated in the form:

(*) What are we implicitly committed to in accepting a theory S and what can we justifiably accept?

As is well-known, (*) has its roots in the work by Kreisel (1958),Kreisel (1970).[2]

Our aim is to consider two possible routes towards a solution, as given over the years by the late Solomon Feferman, starting already in the Sixties and the early Seventies with the work on Predicative Analysis (Feferman (1968)), thereafter with the investigation of self-referential truth in Feferman (1991), and eventually fully transformed by Feferman himself together with Thomas Strahm in Feferman and Strahm (2000).

While the first route – *reflecting* – directly leads into the land of *truth* theories, the second one – *unfolding*, see section 2 – is more mathematical in spirit and hinges upon a point of view, which drives us to the very notion of *operation*.

Andrea Cantini
Department of Letters and Philosophy, via della Pergola 58-60, e-mail: andrea.cantini@unifi.it

[1] *Meditationes de cognitione, veritate et ideis*,1684.

[2] As rightly noticed by a referee, the problem has a much broader and longer history. But we here do not aim to a survey or a historical appreciation of the topics.

Our presentation consists of a survey of the two alternatives, while the implicit commitment issue is specifically dealt within the final section.[3] [4]

Keywords: operation, truth, unfolding, implicit commitment.

2010 Mathematics Subject Classification: 03F03, 03F25, 03F35, 3F40, 03A05.

1 On reflective closure

Let us consider Gödel (1990), p.151, vol II: why is a formalism incomplete? Well, contemplating a *fixed* system gives rise to new axioms, which are as evident as those we start with. But the extension process goes on forever and *can be iterated into the transfinite*. Hence, if we reflect upon a system, we are apparently left with an *infinite task*, and involved or committed to a number of statements which are to be made explicit, at least partially and potentially.

Are we then forced to use the so-called *ordinals*? In the last century this choice has been implemented in several directions, e.g. along the lines of the so-called *ordinal logics* in the sense of Turing, *predicative mathematics*, or else *progressions of theories*; and some conceptual as well as technical difficulties arise, that we list below.

I Ordinals are *external* notions, and we have to cope with a *metabasis eis allo genos*. The difficulty shows that there are abstract higher type notions as required by incompletenss.
II We have to justify the view that ordinals or wellorderings are implicit in the given body of, say, arithmetical knowledge. This is technically much involved, as it essentially hinges upon the use of complex and abstract notions (higher notation systems, collapsing functions, etc.).
III We have to pass from *closed* systems to *open-ended* systems. But how to express in a definite fashion this open-endedness?

Of course, it is impossible to explicitly and completely address I-III within the boundaries of this paper. We can nonetheless make an attempt to fix some points at least informally and intuitively. To this aim, let us remind that ordinals are *generalizations of numbers*, which were created by Cantor in order to control suitable topological processes, and in fact they are *reifications*[5] of *iteration procedures*, that go definitely beyond the realm of *finite* numbers into the so-called *transfinite*. Their role has become essential for assigning *invariants* to computations and

[3] The proposals we consider are extensively and fully developed in Feferman (1991), Feferman and Strahm (2000), Feferman and Strahm (2010), Buchholtz (2013), Buchholtz et al. (2016); as to their connections with the general theme of abstraction, they can also be supplemented by means of results in Cantini (1989), Cantini (1996), Cantini (2016).

[4] This work was supported by the Italian Ministry of Education, University and Research through the PRIN 2017 Program The Manifest Image and the Scientic Image (Prot.2017ZNWW7F004

[5] If a traditional philosophical terminology is here allowed.

proof-theoretic investigations, which involve *consistency*, as well as the *structure of proofs*. In general, one has to cope with order-theoretic structures, which in the standard cases must be *well-founded* (= no infinite regress is possible) and yet can be represented by elementary means, e.g. notation systems, i.e. *symbolic structures* (Leibniz). Moreover ordinals have been crucial – since the Thirties with Gentzen, and later with Schũtte's school, Takeuti, the Bern school, Feferman, the investigations of Girard, Arai, Rathjen's " art of ordinal analysis" Rathjen (2006), etc., till most recent contributions Freund and Rathjen (2021).

1.1 The Kripke–Feferman theory KF

The so-called theory of reflective emerges from the attempt to (partially) overcome difficulty I above. More precisely, if we try to make these ideas formally precise, we are naturally driven to formalize the clauses of Kleene's strong three valued semantical schema for *self-referential truth*, and hence to the well known KF-theory (over PA, as given in Feferman (1991)). For the reader's sake, let us summarize the essential points about KF.

Firstly, as a starting intuitive point, Kreisel (1970) suggested to study *principles of proof and definition that are recognized as valid, once we have understood certain given concepts*: the typical example was predicativism, as *the conception unfolding what is implicit in accepting the notion of natural number*. Formally, this view leads to articulating some form of self-reflection that developed along the so-called *autonomous transfinite progressions* of theories in the Sixties (Feferman (1968)); which is not so natural and technically involved, as it has to deal with well-foundedness by means of elementary methods.

If one avoids transfinite iterations along well-orderings, two routes have been opened (see below and 2).

The first route is to develop a direct *finite ordinal-free alternative* via the notion of *reflective closure* of a system, with its full-schematic variant. Let us recall the preliminary axioms for the simple non-schematic reflective closure. For the sake of a smoother formalization, we adopt some abbreviations.

- In general, we follow the conventional notations of (Halbach, 2011, pp.32-33); hence we use the underscore-dot notation for naming expressions defining function symbols corresponding to logical operations and predicates of our formal language (e.g. $\underset{\cdot}{T}$, $\underset{\cdot}{\neg}$, $\underset{\cdot}{\forall}$, $\underset{\cdot}{\wedge}$, etc.);
- $F(s)$ is a shortening for $T(\underset{\cdot}{\neg}x)$;
- s° stands for the *arithmetical value* of the term s (and this can be defined in the arithmetical language of Peano Arithmetic, see (Halbach, 2011, p.32)).

First of all, the language of reflective closure over PA (alsoknown as KF) expands the language of PA with a monadic predicate T for truth, and $F(s)$ is a shortening for $T(\underset{\cdot}{\neg}x)$. The axioms of KF comprise PA, full induction schema, and

(i). $\forall s\,\forall t\,\Big((T(s\dot{=}t) \leftrightarrow s^\circ = t^\circ)\big) \wedge \big(F(s\dot{=}t) \leftrightarrow s^\circ \neq t^\circ)\Big)$, and similarly for other predicates other than =, except for the special predicates T and F;
(ii). $\forall s\,\big((T(\dot{T}s) \leftrightarrow T(s^\circ)) \wedge (F(\dot{T}s) \leftrightarrow F(s^\circ))\big)$;
(iii). $\forall s\,\big((T(\dot{F}s) \leftrightarrow F(s^\circ)) \wedge (F(\dot{F}s) \leftrightarrow T(s^\circ))\big)$;
(iv). $\forall x\,\big(\mathsf{Sent}_{\mathsf{KF}}(x) \rightarrow (T(\dot{\neg}x) \leftrightarrow F(x)) \wedge (F(\dot{\neg}x) \leftrightarrow T(x))\big)$;
(v). $\forall x\,\forall y\,\big(\mathsf{Sent}_{\mathsf{KF}}(x\dot{\wedge}y) \rightarrow (T(x\dot{\wedge}y) \leftrightarrow T(x) \wedge T(y)) \wedge (F(x\dot{\wedge}y) \leftrightarrow F(x) \vee F(y))\big)$;
(vi). $\forall v\,\forall x\,\big(\mathsf{Sent}_{\mathsf{KF}}(\dot{\forall}vx) \rightarrow (T(\dot{\forall}vx) \leftrightarrow \forall t\,T(x[t/v])) \wedge (F(\dot{\forall}vx) \leftrightarrow \exists t\,F(x[t/v]))\big)$.

Incidentally, it is worth mentioning that KF is intimately related to a logical *development of non-extensional concepts* (classification, operation) and *semantical investigations*; see Cantini (1996) for the connections with Aczel's Frege structures and explicit mathematics. Furthermore, KF has an interesting model theory (rich lattice theoretical results); it is also related to the standard fixed point theory $\widehat{\mathsf{ID}_1}$ (see Feferman (1982)); in turn, this leads towards the foundations of intuitionistic type theory (see Hancock's conjecture, again Feferman (1982)). KF can also be seen as a generalization of the standard theory of compositional truth, and, more naturally, as a generalization of a theory of a positive inductive definition of truth and falsity (see Halbach (2011), §8.7).

1.2 Basic results on KF

For the reader's sake, we add a reminder on significant fragments and their proof theory.

(i) $\mathsf{KF}\lceil$ is KF with induction for T-free formulas;
(ii) KF_c is KF with induction restricted to total predicates, i.e. such that $\forall x T(a(x) \vee T(\dot{\neg}a(x)))$;
(iii) KF_p is KF with induction restricted to internal predicates.

KF and its variants can be compared with subsystems of second order arithmetic, see e.g. Simpson's monograph (Simpson (1999)). Indeed, let $(\Pi^0_1-\mathrm{CA})_{<\lambda}$ stand for (an axiomatic version of) ramified analysis up to any level $<\lambda$ (λ being a constructive ordinal). Let CONS be the formal statement corresponding to the consistency of the truth predicate, i.e., to

$$\forall s(\neg(T(s) \wedge T(\dot{\neg}s))$$

Then:

Theorem 1

(i) $\mathsf{PA} \equiv \mathsf{KF}\lceil + \mathsf{CONS} \equiv \mathsf{KF}_c$;
(ii) $\mathsf{KF}\lceil$ has non-elementary speed-up over PA (see corollary 5.13 of Fischer (2014));[6]
(iii) $(\Pi^0_1-\mathrm{CA})_{<\omega^\omega} \equiv \mathsf{KF}_p + \mathsf{CONS}$;

[6] It is open if $\mathsf{KF}\lceil$ has non-elementary speed-up over its positive compositional fragments.

(iv) $(\Pi^0_1 - \mathrm{CA})_{<\varepsilon_0} \equiv \mathsf{KF} + \mathsf{CONS}.$

For the proof, see Cantini (1989).

The theorem is paradigmatic, in order to grasp at least some flavour of this research. There the function played by the ordinals is made transparent: ordinals are intended to classify truth-theoretic principles and to compare their strength. Infact, they measure how far we can proceed to iterate the arithmetical comprehension principle, which grants the existence of sets, e.g. up to ω^ω or higher up to ϵ_0.

Remark 1 As to the notion of *speed-up* and its precise definition, see (Fischer, 2014, definition 2.3). This technical notion can be informally clarified by the intuition that, as already discovered by Gödel (1936) in the Thirties, adopting abstract notions and principles thereof – like truth or set – usually provide *sensible reduction in length* for (an infinite number of) already available proofs. More explicitly, there are arithmetical theorems, which are provable in $\mathsf{KF}\lceil$ with derivations, which are *much shorter* than those that can be found to exist in PA. [7]

1.3 Schematic reflective closure

Objection III of section 1 above can be overcome with the idea that a schema does not apply to a *fixed language*, but *to any language which one comes to recognize as embodying meaningful basic notions* (see p. 189, Strahm (2017)). Informally, this suggests a rule which allows to make inferences from schemata accepted in the *original arithmetical language* to schemata of the *full* language. In particular, we can extend KF *by a suitable substitution rule* of the form

$$\frac{\varphi(P)}{\varphi(\hat{x}\psi(x))}$$

where φ is a formula of $\mathcal{L}_{\mathsf{PA}}$, the language of Peano arithmetic with an additional predicate symbol P, ψ is arbitrary (hence T, P can occur in ψ). $\varphi(\hat{x}\psi(x))$ stands as an abbreviation for the formula resulting from $\varphi(P)$ when each occurrence of a subformula $P(t)$ in $\varphi(P)$ is replaced by $\psi(t)$.[8]

The resulting system yields the *schematic reflective closure* $\mathit{Ref}^*(\mathsf{PA}(P))$, another way to characterize predicativity in the sense of Feferman and Schütte:

Theorem 2 (Feferman (1991))

$$\mathit{Ref}^*(\mathsf{PA}(P)) \equiv (\Pi^0_1 - \mathrm{CA})_{<\Gamma_0}$$

[7] I.e. if m is the length of a proof of A in $\mathsf{KF}\lceil$ then the length of a proof p of A in PA might require many iterations of the exponential $exp_2(x) = 2^x$, i.e. a tower $exp_2(exp_2(\ldots exp_2(m)))\ldots)$, in order to find an upper bound in terms of m to the length of a purely arithmetical proof of A in PA.

[8] $\hat{x}\psi(x)$ is the classical Russell-Whitehead notation for the class defined by $\psi(x)$; it is assumed that substitutions are correct, i.e. no confusion of variables can occur.

Here Γ_0 is the well-known Feferman-Schütte upper bound of the predicative ordinals, and hence of those inductive methods that can be accepted predicatively and rely upon reflection and reflective closure. Is it possible to fix more structural routes? The answer is positive, as seen from the next section.

2 Feferman-Strahm's unfolding

Objections can be raised against reflective closures of theories (Strahm (2017) p.189): Feferman (2016), p. 282) states that *the resulting theories have still an air of artificiality*. More specifically:

(1) it is questionable whether the truth axioms have the same evidence as the ground axioms without truth;
(2) the formalization of syntax is an essential ingredient together with the coding machinery, and this somewhat conceals the ontological nature. Indeed, it tends to conceal the conceptual depth, yielding an air of taxonomy and mere linguistic reality.

It follows that, as stated in the opening abstract, a second operational route to the analysis of implicit commitment deserves due attention: one has to cope with (1)-(2), and the point is to attain *a syntax-free formalization* of the main notions. In other words, in order to understand the truth axioms, you assume suitable operations – mainly on numbers and finitary structures such as syntactical expressions, but also quantifiers – which ought to be made explicit as *direct objects of our reflecting procedure*.

2.1 Unfolding informally presented

The idea is that, given a system S, *unfolding* S means to define a theoretical framework, where

- we have operations and predicates *understood as partial operations*;
- *logical* operations on propositions and predicates are given;
- operations may be applied to operations, so that self-application is not forbidden, i.e. a theory of untyped operations is assumed;
- partial recursive operations can be regarded as a possible model of our operations and hence non-extensionality holds.

Indeed, it turns out that the system of operations is at least as expressive as untyped lambda calculus and combinatory logic; and this step is a natural way to achieve independence *on encoding*. More explicitly, one assumes to live in an applicative structure, indeed a *partial combinatory algebra*, whose language consists of:

- application terms ts for expressing the application of t to s, $t\downarrow$ for expressing that t converges; application is *strict*, i.e. if ts is defined then both t, s are defined;
- a logic of partial terms and the relation $t \simeq s$ for expressing that $t = s$, whenever t, s are both defined;
- the so-called basic combinators K,S with standard axioms $Kts \simeq t$ and $Stsr \simeq tr(sr)$, whence λ abstraction and recursion operators become available; the system of operations is also equipped by pairing and projections, definition by cases.

For the reader's sake, we like to stress the informal reading, i.e. the natural effective interpretation of $xy \simeq z$ as: *the algorithm repesented by x halts (i.e. converges) on the input y providing z as output*. For details, see (Eberhard and Strahm, 2015, pp. 158–160).

2.2 Unfolding axiomatized

Let me now describe a few details for the unfolding of NFA – an acronym for *non-finitist arithmetic* –, which is regarded as the paradigmatic *ur*-system. The axioms of NFA itself are simply the usual ones for 0, sc (successor) and pd (predecessor), together with the *induction axiom*, given as $P(0) \wedge \forall x[P(x) \rightarrow P(sc(x))] \rightarrow \forall x(P(x))$, where P is a free predicate variable. The language of the unfolding of NFA adds a number of constants, the predicate symbol $N(x)$ (x is natural number), the predicate symbol $\Pi(x)$ (= 'x is a predicate), and the intensional membership relation $y \in x$ for x such that $\Pi(x)$.

Then the proper axioms of *the unfolding* $\mathsf{U}(\mathsf{NFA})$ *for non-finitist arithmetic* consist of the following groups[9] :

(I) the axioms of NFA relativized to N, the collection of natural numbers;
(II) the partial combinatory axioms, with pairing, projections and definition by cases;
(III) an axiom for the characteristic function of equality on N;
(IV) axioms for various constants in the domain Π of predicates, namely for the natural numbers, equality, the free predicate variable P, and for the logical operations $\neg$, $\wedge$, $\forall$ and inverse image of f along a; [10]
(V) an axiom (*join*)) for the *disjoint union* $\mathsf{j(f)}$ of a sequence f of predicates over numbers: whenever $f : N \rightarrow \Pi$, $\mathsf{j(f)}$ is the collection of all ordered pairs (u, v) where $u \in f(v)$, v being a natural number.

In analogy with the case of schematic reflective closure, *the full unfolding* $\mathsf{U}(\mathsf{NFA})$ is then obtained by applying the *substitution rule* $A(P)/A(B)$, where B is an *arbitrary formula* of the unfolding language.

---

[9] For formal definitions and details, see Strahm (2017).

[10] The inverse image of an operation f along a predicate a is - extensionally - the collection of all x such that $fx \in a$.

The *operational unfolding* $\mathsf{U}_0(\mathsf{NFA})$) is simply obtained by restricting to axiom groups (I)-(III) and with the formulas B in the substitution rule restricted accordingly. In $\mathsf{U}_0(\mathsf{NFA})$ one successively constructs terms $t(x)$ intended to represent each primitive recursive function, by means of the recursion operator and definition by cases. By applying the substitution rule, it is then shown by induction on the formula $t(x)\downarrow$ that each such term defines a total operation on the natural numbers. Thus the language of PA may be interpreted in that of $\mathsf{U}_0(\mathsf{NFA})$ and hence – by application of the substitution rule once more – we have that PA itself included in that system.

In $\mathsf{U}(\mathsf{NFA})$ the domain of predicates is considerably expanded by use of the *join operation*. Once we have established that a primitive recursive ordering $<$ satisfies the schematic transfinite induction principle $TI(<, P)$ with the free predicate variable P, we can apply the substitution rule, in order to carry out proofs by induction on $<$ with respect to *arbitrary formulas*. In particular, one can establish the existence of a predicate corresponding to the the so-called hyperarithmetical hierarchy[11] along such an ordering, relative to any given predicate p in Π, as naturally axiomatized by iterating arithmetical comprehension. Then by means of the usual arguments, if one has established in $\mathsf{U}(\mathsf{NFA})$ the schematic principle of transfinite induction along a standard ordering for an ordinal α, one can lift it upwards to a wellordering of type $\varphi\alpha 0$,[12] and hence the same for each ordinal less than Γ_0, the upper bound for predicative reasoning. Thus $\mathsf{U}(\mathsf{NFA})$ can deal with the ramified analytic systems up to Γ_0 and it essentially matches up with the theory of theorem 2.

The main results of Feferman and Strahm (2000) are that (i) $\mathsf{U}_0(\mathsf{NFA})$ is proof-theoretically equivalent to PA and is conservative over it; (ii) $\mathsf{U}(\mathsf{NFA})$ is proof-theoretically equivalent to the union of the ramified analytic systems up to Γ_0 and is conservative over it.

In other words, $\mathsf{U}(\mathsf{NFA})$ is proof-theoretically equivalent to *predicative analysis* as charaterized in theorem 2. In addition, the intermediate system $U_1(NFA)$ without the join axiom (V) is proof-theoretically equivalent to the union of the ramified systems of *finite* level. While the unfolding of non-finitist arithmetic attains the limits of predicativism, it is proper to the essence of predicativism that *ascent towards more complex sets is only allowed through restricted quantification, i.e. quantifications over totalities that can be regarded as already given*, and hence only *predicative ordinals* are accepted, i.e. ordinals which can be recognized by exclusive appeal to notions that have already been secured and are generated from below, bottom-up.

Indeed, these ideas have a more general import: constructive theories – Martin-Löf type theory, and Myhill and Aczel set theory – *feature predicativity as a distinctive form of constructivity* (see Crosilla (2017), Crosilla (2018)). This may be condensed into a *constructibility* requirement for sets, which ought to be finitely specifiable in terms of uncontroversial primitive objects and simple operations over them. Historically, predicativity emerged at the beginning of the 20th century as a component of an influential analysis of the paradoxes by Poincaré and Russell. According to this analysis, the paradoxes are caused by vicious circles in definitions; and adherence

[11] In essence a version of the ramified hierarchy in sense of Russell which is iterated up to the first non-recursive ordinal.

[12] $\varphi\alpha\beta$ is the so-called Veblen hierarchy; for details, see Feferman (1991) , Rathjen (2006).

to predicativity was therefore proposed as a systematic method for preventing such problematic circularity.

Now a *subtler role* of ordinals occurs in the case of *impredicative* definitions, which can be analyzed up to a certain point via ordinal-theoretic methods, but with a *distinctive* feature: the unfolding of certain definitions and processes dealing with finite objects, e.g. proofs, requires to climb up to and to compute with certain ordinals, which *surpass the first uncountable ordinal*. But then we have to collapse down onto a *countable* ordinal, in order to recover information on usual proofs and definitions! Typically – as in the case of the so called ordinal $\psi(\Gamma_{\Omega+1})$ defined in Buchholtz (2013), Buchholtz et al. (2016), which parallels the ordinal Γ_0 of predicative analysis – one defines a *countable* ordinal, which necessarily requires – to be defined – the *uncountable* ordinal Ω and $\psi(\Gamma_{\Omega+1}) < \Omega$.

Incidentally, let me conclude by mentioning that these investigations contribute to the research program of *metapredicativit*y (Jäger and the Bern school; see Ranzi and Strahm (2019)). The idea is that, while predicative methods build up objects from below, impredicative methods assume the existence of large objects in order to construct smaller ones. In the case of metapredicativity, one generalizes methods from predicative proof theory to investigate theories that are impredicative in the sense of Feferman and Schütte.

2.2.1 Unfolding classified

It is interesting to look for generalizations, and hence to define the unfolding $\mathsf{U}(\mathsf{S})$ for given S. Indeed, for special choices, *significant theories* arise. Just to give concrete instances of this sort of phenomena, let us choose for S two standard theories: an axiomatization of *Polynomial Time Arithmetic* PTCA, as given, say, by Ferreira (1990) or Eberhard and Strahm (2015), and PRA, a standard version of *Primitive Recursive Arithmetic*. Now it is also known that the two systems can be rephrased as *applicative theories, as based on the notion of self-applicable operation,* and yet their computable content (the algorithms recognized as well-defined in the two systems) coincide – respectively – with the polytime operations and the primitive recursive operations. In other words, we are led to consider:

- FEA, a version of feasible arithmetic as defined by Eberhard and Strahm (2015)[pp.156–57]. Roughly, FEA is an axiomatization of partial combinatory logic together with an underlying structure: the theory of binary strings (finite sequences of bits), as endowed with the natural operations of concatenation and string product and with a relation $\tau \leq \sigma$, meaning that the length of the string τ is less than the string σ.
- FA, the basic system of *finitist arithmetic*, which includes a basic theory of operations as based on partial application with the basic constructions for producing constant operations and substitutions, and includes standard partial combinatory logic, succcessor, predecessor, pairing operations and projections. As is well-known, partial combinatory logic allows *self-application*; in order to understand it, it is enough to follow the informal reading $xy \simeq z$ as: *the algorithm coded by x*

applied to the input y converges and produces z as output. Besides application, the intended universe includes natural numbers; hence three basic statements are possible: $N(t)$, $t = s$, $ts \simeq r$.
The essential point is that the logical operations now are *restricted* to $\wedge$, $\vee$, $\exists$ existential quantifiers. Provable propositions $A(x)$ are interpreted as verifying $A(n)$ for each natural number n, but *we do not have universal quantification* over the natural numbers as a logical operation. Nor do we have negation (except of numerical equations), which, when applied to existential formulas, could be interpreted as having the effect of universal quantification.

- $\mathsf{U(FEA)} = \mathsf{PTCA}$ (Eberhard and Strahm (2015)); this is related to *feasibility*, and only *bounded quantification on binary strings* is allowed, while one has a weak notion of truth closed under bounded quantification, $\wedge$, $\vee$, positive prime conditions;
- $\mathsf{U(FA)} = \mathsf{PRA}$; this corresponds to *finitism* in the sense of Tait (Feferman and Strahm (2000));
- Let BR be the so-called *bar rule*, i.e. the inference which roughly states: if $<$ is a well-founded ordering relation – in symbols $\mathsf{WF}(<)$ –, then each instance of the transfinite induction schema $\mathsf{TI}(<, \mathsf{B})$ on $<$ is allowed (B arbitrary formula of the full language). Then $\mathsf{U(FA + BR)} = \mathsf{PA}$; this is inspired by Kreisel's idea that PA is the *least upper limit* of finitism, as characterized by means of Gentzen's consistency proof via transfinite induction on a suitable natural well-orderings;
- $\mathsf{U(ID_1)}$, the unfolding of the elementary theory of inductive definitions, corresponds to overcome the limits of standard predicativity towards metapredicativity and beyond; this yields is a theory of strength $\psi(\Gamma_{\Omega+1})$, but other equivalent systems can be found in Buchholtz et al. (2016).

Feferman and Strahm (2000), Feferman and Strahm (2010), Buchholtz et al. (2016) develop a suitable proof-theoretic analysis of the systems.

Let us mention an alternative strategy to the proof-theoretic results given in Feferman and Strahm (2000), Feferman and Strahm (2010): they can be easily obtained by interpreting the unfolding systems into theories of *abstract truth over combinatory structures*, as outlined in Cantini (2016), Cantini (1996), thus establishing another bridge with *appplicative systems*.

More explicitly, let PT be a theory of propositions and truth, which, roughly, consists of (i) the axioms for combinatory logic enriched by numbers; (ii) natural compositional axioms for truth T and propositions P; (ii) the schema of number theoretic induction (for details, see Cantini (2016)).

Now in PT the collections of propositional functions can simulate a rather rich *structure of types*; in particular it is closed under elementary comprehension and join, and it turns out:

Theorem 3 $\mathsf{U(NFA)}$ without substitution rule is interpretable into the system PT.

Remark 2 As to the import of $\mathsf{U(NFA)}$, observe that PT without substitution is known to have the same upper bound as ramified analysis of any level below ϵ_0 (ϵ_0 being the well-known proof theoretic ordinal discovered by Gentzen in the thirties and associated to the consistency of PA). Of course, this means that U(NFA) is a highly non-trivial system and worth considering.

Remark 3 The theories of propositions and truth we just cited are in a sense intermediate betweem reflective closures and unfolding, as they fully integrate the operational side, while making sense of an abstract notion of truth.

3 On the Implicit Commitment Thesis ICT

The implicit commitment thesis ICT (Dean, 2014, p.32) states that *in accepting a formal systems S one is also committed to additional resources not available in the starting theory S but whose acceptance is implicit in the acceptance of S.* Of course, this is only an informal statement and one might want to look for a more definite form of it. For instance, if we accept the axioms of a theory S together with logical inferences and principles, we can accept as basic the notion of proof in S. Eventually, we are led to accept formal proofs in S as sound (whence the formal consistency of S), which formally corresponds to the acceptance of a *reflection principle* RFN(S) for S:

- if S proves a statement φ, then φ is true.

As to RFN(S), either we regard it as a schema in the given formula φ, or else we explicitly add a truth predicate (see the next subsection).

Recently, this step has been questioned by Dean (2014), essentially on the ground of the following argument: certain systems – incompleteness notwithstanding – are plausibly *complete* with respect to a given body of knowledge, typically PRA for *finitism* in the sense of Tait (1981), or PA with respect to *genuine 1st order number theoretic properties* (this is known as Isaacson's thesis, see Isaacson (1987)). Here by *genuine 1st order* it is meant that there are *no hidden higher order notions* around: or, put differently, the content is essentially arithmetical, no reference to the *notion of set* of natural numbers.

But, e.g., if we accept Isaacson's thesis IT, we *cannot be committed* to a reflection principle RFN(PA) as implicit, since it goes far beyond the boundaries of PA, and hence the principle makes PA *unstable*, in the sense that it is naturally amendable, and also incomplete, against the thesis. Hence this fact conflicts with ICT; and it seems that we have to give up ICT or to modify ICT in some sense.

Furthermore ICT apparently yields *ontological* consequences. For instance, *consistency*, which follows from ICT, implies the existence of structures and models by the so-called arithmetized completeness theorem (see Dean (2020)), and this fact apparently conflicts with the idea that truth is accepted as a *thin* notion, according to deflationism Cieslinski (2017). Hence assuming ICT conflicts with deflationism. i.e. the thesis that truth is thin and unsubstantial.

3.1 ICT sharpened: semantical way-out

Let us try a fresh start. Taking ICT seriously leads to extend the ground system S with a notion of *truth*. Just to be concrete, choose S to be PA. Accepting PA, we are then committed to a theory of arithmetical truth CT(PA), which embodies the *standard compositional clauses*. Essentially, CT(PA) embodies the axioms of KF *restricted to the codes of* PA*-formulas*. Therefore, if we also allow full induction – namely also applied to formulas where T occurs –, we can prove the reflection schema (and hence the consistency) of PA in the form:

$$\forall x(Sent_S(x) \wedge Axiom_S(x) \rightarrow T(x))$$

But, if we accept the criticism above, CT(PA) is *inadequate*. We simply ought to consider weaker systems with restricted induction, e.g. CT(PA)⌈, i.e. the extension of PA in the truth-theoretic language, as given above, which contains

1. the *compositional* axioms for truth, as applied to (codes of) PA-sentences;
2. the axiom stating that $\forall x(For_{\mathsf{PA}}(x) \rightarrow TInd(x)$ where $Ind(x)$ formalizes that x is the code of an arithmetical instance of the induction schema (for details see Halbach (2011)).

At this stage, it turns out that we can apply non-trivial results in Kotlarski et al. (1981), Leigh (2015), Enayat and Visser (2015) about theories endowed with a truth predicate, implying that

Theorem 4 CT(PA)⌈ has the same arithmetical content as PA.

Recently, the theorem has triggered a new proposal, put forward by Nicolai and Piazza (2019). In essence, it amounts to state that the implicit commitment IC has a composite nature, a *variable* component and an *invariable* one, the so-called *semantic core*, i.e. a set of of semantical principles about truth we are naturally committed to accept, and yet these principles are conservative over the ground system. More explicitly, the semantic core should include, besides the compositional truth axioms for PA-sentences, the axioms stating the truth of all propositional tautologies, the fact that the inference rules (modus ponens is enough) preserve truth, the truth of all its non-logical axioms.

However, it is not clear yet whether the theorem above can be strengthened to the effect that the conservation result keeps holding for the extended system (compare with Nicolai and Piazza (2019), p.929, footnote.).

Lastly, let us mention that another route is viable via KF, i.e. we can propose a notion of *reflective* semantic core. This means that we can regard truth as self-referential truth, and hence subscribe the axioms in KF, including the axiom $LogT$ formalizing the statement: all logical PA-axioms in the language of PA are true. This can easily be stated in a standard formalization of the syntax of PA:

- $\forall x(Ax_{PA}(x) \rightarrow T(x))$

Of course, we may choose $\mathsf{KF_c}$ as a variant of semantic core, i.e. a sort of reflective semantic core, i.e. $\mathsf{KF_c}$. However, adopting $\mathsf{KF_c}$ is not really a *panacea*: there remain problems similar to those of $\mathsf{CT}(\mathsf{PA})\lceil$, which are left open as well, e.g. it remains to be seen if $\mathsf{KF_c}$ with *LogT* is still conservative over PA.

And a crucial point ought to be clarified though: once truth as a resource is introduced, the reflection process should be applied to induction instances applied for *total* conditions. Why should it be so?

3.2 Ontological way-out

Assume we consider the unfolding approach: can we avoid the previous difficulties? Well, we can rephrase ICT as an attempt of extracting genuine mathematical notions. Truth is no more accepted as a starting point, against a metatheoretical approach, but *operations* form the basic constituents.

Of course, operations are understood intensionally, as rules, and not defined via set theoretical methods. Operations should be regarded as acting on the intended universe of individuals of the underlying theory, as well as on a domain of predicates and operations themselves. Both domains are included in a comprehensive domain V and the closure conditions ensure, as already stated, that we have a ground applicative structure, typically a partial combinatory algebra.

It also turns out that, as Strahm puts it, the basic aim for implicit commitment becomes wider in scope: given a schematic system S, *which operations and predicates, and which principles* concerning them ought to be accepted if one has accepted S? Eventually, elaborating the unfolding program directly leads to a confluence within the framework of the so-called applicative frameworks, that have been developed at length in the context of the so-called program of explicit mathematics (see section 2 in Feferman (2016)).

4 Conclusion

Let us recall ICT, as stated by Dean (2014), p.32, in the form of a thesis:

> *Anyone who accepts the axioms of a mathematical theory S is thereby also committed to accepting various additional statements* Δ *which are expressible in the language of S but which are formally independent of its axioms.*

We wonder whether there might be a sensible restatement of Dean's ICT. We observed that *unfolding* induces a switch away from reflection principles: the problem is not quite to guess suitable reflection principles – as schemata or axioms – but to look for *new operations* or *predicates*, and corresponding *closure properties*, possibly in search of new tangible, significant features of the objects involved.

Hence we depart from the issue of implicit commitment and from a problem involving *the limits of the axiomatic method*, and we are moving to a problem of *mathematical content*. What is the effect of reflection principles on the kind of entities a given theory is calling into existence? In order to handle increasing logical complexity, we are pushed towards possibly higher type operations, more than stronger and stronger kind of inferences rules: think of the use of notion of universe in type theory (Martin-Löf, Feferman) and the emerging of abstract objects which concern proofs but are not merely logico-linguistic, as already envisaged in several contexts (Hilbert himself, Schütte and the Münich school, Girard).

As to the traditional ideas underlying ICT, and in order to make clear the open character of this note, we like to conclude with a few open questions: is there any objectual counterpart of the conceptual frame supporting a reflection principle somewhat implicit in accepting the axioms of PA? Is there a structural content of the idea of semantic core? What is the use of the semantic core? Does it help in making proofs more perspicuous or elegant or shorter?

References

Buchholtz, U. (2013). *Unfolding of systems of inductive definitions*. Ph.d. thesis, Stanford University.

Buchholtz, U., G. Jäger, and T. Strahm (2016). Theories of proof-theoretic strength $\psi(\gamma_{\epsilon_{\Omega+1}})$. In *Concepts of proof in mathematics, philosophy, and computer science*, pp. 115–140. Berlin: De Gruyter.

Cantini, A. (1989). Notes on formal theories of truth. *Z. Math. Logik Grundlag. Math. 35*, 97–130.

Cantini, A. (1996). *Logical Frameworks for Truth and Abstraction*. Amsterdam: North Holland.

Cantini, A. (2016). About truth and types. In R.Kahle, T. Strahm, and T. Studer (Eds.), *Advances in proof theory*, pp. 31–64. Cham: Birkhäuser-Springer.

Cieslinski, C. (2017). *The Epistemic Lightness of Truth. Deflationism and its Logic*. Cambridge: Cambridge University Press.

Crosilla, L. (2017). Predicativity and Feferman. In G. Jäger and W. Sieg (Eds.), *Feferman on Foundations*, Volume 17 of *Outstanding Contributions to Logic*, pp. 423–447. Cham: Springer.

Crosilla, L. (2018). Exploring predicativity. In H. K. Mainzer, P.Schuster (Ed.), *Proof and Computation. Digitization in Mathematics, Computer Science and Philosophy*, pp. 83–108. Singapore: World Scientific Publishing.

Dean, W. (2014). Reflection and the provability of soundness. *Philos. Math. 23*, 31–64.

Dean, W. (2020). Incompleteness via paradox and completeness. *Rev. Symb. Log. 13*, 541–592.

Eberhard, S. and T. Strahm (2015). Unfolding feasible arithmetic and weak truth. In T. Achourioti, H. Galinon, J. Martinez, and F. F. Fujimoto (Eds.), *Unifying the philosophy of truth*, pp. 153–167. Cham: Springer.

Enayat, A. and A. Visser (2015). New constructions of satisfaction classes. In T. Achourioti, H.Galinon, J.M.Martinez, and K. Fujimoto (Eds.), *Unifying the philosophy of truth*, pp. 321–35. Cham: Springer.

Feferman, S. (1968). Autonomous transfinite progressions and the extent of predicative mathematics. In *Logic, Methodology and Philos. Sci. III*, Amsterdam, pp. 121–135. North-Holland.

Feferman, S. (1982). Iterated inductive fixed-point theories: application to Hancock's conjecture. In *Patras Logic Symposion*, Amsterdam, pp. 171–196. North-Holland.

Feferman, S. (1991). Reflecting on incompletenss. *Journal of Symbolic Logic 56*, 1–49.

Feferman, S. (2016). The operational perspective: three routes. In R. Kahle, T. Strahm, and T. Studer (Eds.), *Advances in proof theory*, pp. 269–289. Cham: Birkhäuser-Springer.

Feferman, S. and T. Strahm (2000). Unfolding finitist arithmetic. *Ann.Pure Appl. Logic. 104*, 75–96.

Feferman, S. and T. Strahm (2010). The unfolding of finitist arithmetic. *Rev. Symb.Log. 3*, 665–689.

Ferreira, F. (1990). Polynomial time computable arithmetic. In *Logic and computation*, pp. 137–156. Providence, RI: Amer. Math. Soc.

Fischer, M. (2014). Truth and speed-up. *Rev. Symb.Log. 7*, 319–340.

Freund, A. and M. Rathjen (2021). Derivations of normal functions in reverse mathematics. *Ann. Pure Appl. Logic 172*, 1–49.

Gödel, K. (1936). Über die Länge von Beweisen. In K. Menger (Ed.), *Ergebnisse eines Mathematischen Kolloquiums*, Volume 7, pp. 23–24.

Gödel, K. (1990). Remarks before the princeton bicentennial conference on problems in mathematics. In S. e. a. Feferman (Ed.), *Collected Works, vol. II: Publications 1938-1974*, pp. 150–54. Oxford: Oxford University Press.

Halbach, V. (2011). *Axiomatic Theories of Truth* (1st edition ed.). Cambridge: Cambridge University Pres.

Isaacson, D. (1987). Arithmetical truth and hidden higher order concepts. In J. et alii (Ed.), *Logic Colloquium '85*, pp. 147–169. Amsterdam: North-Holland.

Kotlarski, H., S. Krajewski, and A. Lachlan (1981). Construction of satisfaction classes for nonstandard models. *Canad. Math. Bull. 24*, 283–293.

Kreisel, G. (1958). Ordinal logics and the characterization of informal notions of proof. In *Proc. International Congress of Mathematicians, 14-21 August 1958.*, pp. 289–299. Cambridge: Cambridge University Press.

Kreisel, G. (1970). Principles of proof and ordinals implicit in given concepts. In A. Kino, J. Myhill, and R. E. Vesley (Eds.), *Intuitionism and Proof Theory*, pp. 489–516. Amsterdam: North Holland.

Leigh, G. E. (2015). Conservativity for theories of compositional truth via cut elimination. *J. Symb. Log. 80*, 845–865.

Nicolai, C. and M. Piazza (2019). The implicit commitment of arithmetical theories and its semantic core. *Erkenntnis 84*, 913–937.

Ranzi, F. and T. Strahm (2019). A flexible type system for the small veblen ordinal. *Arch. Math.Logic 58*, 711–751.

Rathjen, M. (2006). The art of ordinal analysis. In *International Congress of Mathematicians*, Volume II, pp. 45–69. Zürich: Eur. Math. Soc.

Simpson, S. (1999). *Subsystem of Second order Arithmetic*. Berlin Heidelberg: Allen and Unwin.

Strahm, T. (2017). Unfolding schematic systems. In G. Jäger and W. Sieg (Eds.), *Feferman on foundations*, pp. 187–208. Cham: Springer.

Tait, W. (1981). Finitism. *J. Philos. 78*, 524–546.