Proceedings Paper:
Rathjen, M and Toppel, M (2019) On Relating Theories: Proof-Theoretical Reduction. In: Centrone, S, Negri, S, Sarikaya, D and Schuster, PM, (eds.) Mathesis Universalis, Computability and Proof. Humboldt-Kolleg: Proof Theory as Mathesis Universalis, 24-28 Jul 2017, Loveno di Menaggio (Como), Italy. Springer , pp. 311-331. ISBN 978-3-030-20446-4

https://doi.org/10.1007/978-3-030-20447-1_16

On relating theories: Proof-theoretical reduction

Michael Rathjen and Michael Toppel
Department of Pure Mathematics, University of Leeds
Leeds LS2 9JT, United Kingdom
E-Mail: M.Rathjen@leeds.ac.uk

Abstract

The notion of *proof-theoretical* or *finitistic reduction* of one theory to another has a long tradition. Feferman and Sieg [13, Chap. 1] and Feferman in [22] made first steps to delineate it in more formal terms. The first goal of this paper is to corroborate their view that this notion has the greatest explanatory reach and is superior to others, especially in the context of foundational theories, i.e., theories devised for the purpose of formalizing and presenting various chunks of mathematics.

A second goal is to address a certain puzzlement that was expressed in Feferman's title of his Clermont-Ferrand lectures at the Logic Colloquium 1994: "How is it that finitary proof theory became infinitary?" Hilbert's aim was to use proof theory as a tool in his finitary consistency program to eliminate the *actual infinite* in mathematics from proofs of *real statements*. Beginning in the 1950s, however, proof theory began to employ infinitary methods. Infinitary rules and concepts, such as ordinals, entered the stage.

> *In general, the more that such infinitary methods were employed, the farther did proof theory depart from its initial aims and methods, and the closer did it come instead to ongoing developments in recursion theory, particularly as generalized to admissible sets; in both one makes use of analogues of regular cardinals, as well as "large" cardinals (inaccessible, Mahlo, etc.).* ([19]).

The current paper aims to explain how these infinitary tools, despite appearances to the contrary, can be formalized in an intuitionistic theory that is finitistically reducible to (actually Π^0_2-conservative over) intuitionistic first order arithmetic, also known as Heyting arithmetic. Thus we have a beautiful example of Hilbert's program at work, exemplifying the Hilbertian goal of moving from the ideal to the real by eliminating ideal elements.

Keywords: relative interpretability, partial conservativity, proof-theoretical reduction, infinite proof theory, ordinal analysis
MSC2000: 03F50; 03F25; 03E55; 03B15; 03C70

1 Introduction

Leibniz conceived of mathematics in a very modern way as *mathesis universalis*, i.e., as the most general theory of structures and their relationships. As formal correlates to structures we have axioms abstracted from them that subsequently become organized into axiom systems or theories which describe and classify these structures. Husserl, who called them "Theorieformen", was one of the first thinkers to adumbrate a systematic study of theories as objects of investigation in their own right.[1]

[1] See Centrone [16].

> *Die in solcher Abstraktion definierten Theorienformen lassen sich nun zueinander in Bezug setzen, sie lassen sich systematisch klassifizieren, man kann solche Formen erweitern und verengen, man kann irgendeine vorgegebene Form in systematischen Zusammenhang mit anderen Formen bestimmt definierter Klassen bringen und über ihr Verhältnis wichtige Schlüße ziehen.* [36, p. 431]

The move to the study of formal theories as objects of mathematical investigations in their own right is of course most prominent in Hilbert's *metamathematics* and his new science of *Beweistheorie* (proof theory). This article will be concerned with various relationships that obtain between theories and in particular with notions of reductions between mathematical theories. In the natural sciences, reduction often serves the purpose of explaining the objects and phenomena of one science in terms of more basic or fundamental objects and phenomena of another science, such as chemical reaction in terms of quantum mechanics and heredity in terms of genes. Reductions between mathematical theories are also important in mathematics as for instance in Descartes recasting of Euclidian geometry in analytic geometry and Hilbert's interpretations of non-Euclidian geometries in the theory of real numbers. In this article, however, the main interest lies in foundational theories developed for the purpose of accounting for large parts of mathematics of which first order number theory (also known as Peano Arithmetic, **PA**), second order number theory ($\mathbf{Z}_2$) and Zermelo-Fraenkel set theory with the axiom of choice (**ZFC**) are canonical examples.

There are many important relations between theories that provide reductions of one theory to another: relative interpretability, double negation, functional and realizability translation, (partial) conservativity, and proof-theoretic or finitistic reducibility. In the first part of this article we shall have a closer look at several notions of reducibility and argue that the notion of proof-theoretic reducibility is the most important and most encompassing.

The second part of the paper addresses a fundamental problem in proof theory. In ordinal analysis the proof-theoretic strength of a theory T is often determined by embedding T into an infinitary proof system and showing cut elimination for the infinitary system, where the lengths and cut ranks of derivations are measured by ordinals from an ordinal representation system. How is that that this kind of detour through the infinite still gives rise to finitistic reductions between theories?

2 Reductions

Let's begin with the first and simplest notion. By a theory T, which always comes associated with a language $\mathcal{L}$, we mean a set of sentences of $\mathcal{L}$ (to be thought of as the non-logical axioms of T rather than the set of its theorems). If T is understood we refer by $\mathcal{L}(T)$ to its language.

Given theories T and T', we say that T has a *relative interpretation* in T' if the primitives of $\mathcal{L}(T)$, i.e., the relation, constant, function symbols and the scope of the variables of $\mathcal{L}(T)$ can be defined in $\mathcal{L}(T')$ in such a way that every theorem of T is transformed into a theorem of T' via this translation. In more detail this means the following:

Definition 2.1 1. Let $\mathcal{L}_1$ be a language and T be a theory with language $\mathcal{L}_2$. An interpretation function * of $\mathcal{L}_1$ into T is given by:

(a) An $\mathcal{L}_2$-formula $\chi(x)$ (x being its sole free variable) such that $T \vdash (\exists x)\chi(x)$.[2]

[2] $\{x \mid \chi(x)\}$ constitutes the domain over which the interpreted quantifiers of $\mathcal{L}_1$-formulas will be ranging; so quantifiers become *relativized* to $\{x \mid \chi(x)\}$.

(b) For each constant c of $\mathcal{L}_1$, a formula $\psi_c(x)$, with all free variables exhibited, such that

$$T \vdash \exists! x\,(\chi(x) \wedge \psi_c(x)).$$

(c) For each n-ary predicate P of $\mathcal{L}_1$, a formula $\psi_P(x_1, \ldots, x_n) \in \mathcal{L}_2$ with all free variables exhibited.

(d) For each n-ary function symbol f of $\mathcal{L}_1$ a formula $\psi_f(x_1, \ldots, x_n, y) \in \mathcal{L}_2$, with all free variables exhibited, such that

$$T \vdash \bigwedge_{i=1}^{n} \chi(x_i) \to \exists! y\,(\chi(y) \wedge \psi_f(x_1, \ldots, x_n, y)).$$

The translation * is then lifted to formulas of $\mathcal{L}_1$ in the obvious way. At the atomic level we use the formulas given above for predicate, constant and function symbols respectively. * distributes over the logical particles as expected:

$$\begin{array}{rcl} (\neg\varphi)^* & \equiv & \neg\varphi^* \\ (\varphi \circ \psi)^* & \equiv & \varphi^* \circ \psi^* \text{ for } \circ \in \{\wedge, \vee, \to\} \\ (\exists x\, \varphi(x))^* & \equiv & \exists x\,(\chi(x) \wedge \varphi(x)^*) \\ (\forall x\, \varphi(x))^* & \equiv & \forall x\,(\chi(x) \to \varphi(x)^*). \end{array}$$

2. Assume that S and T are two first-order theories. Then S is said to have a *relative interpretation in T via the function* * (of $\mathcal{L}(S)$ into T) if for every axiom $\varphi(x_1, \ldots, x_n)$ of S,

$$T \vdash \bigwedge_{i=1}^{n} \chi(x_i) \to \varphi^*(x_1, \ldots, x_n).$$

We say that S is *r-interpretable in T* (in symbols $S \trianglelefteq T$) if S has a relative interpretation in T for some translation $*$.

3. If S and T are mutually r-interpretable in each other, we signify this by $S \equiv_{ri} T$.

4. This concept of *relative interpretation* is sometimes credited to Tarski although it was informally and implicitly used before by Hilbert and others. In the more recent usage one omits "relative" (see [40]). However, as there are many flavors of interpretation that we will be discussing, e.g. realizability and functional interpretations, it will be convenient to stick to the name "relative interpretation" to avoid ambiguity.

There are some interesting cases, where the relation $\trianglelefteq$ can be expressed in terms of partial conservativity. The latter notion will be of great importance for this paper, so we turn to it next.

Definition 2.2 We recall the usual stratification of formulas of the language of first-order arithmetic or Peano arithmetic, **PA**, into the classes of Σ^0_k and Π^0_k-formulas. A formula is said to be in Σ^0_k if it is of the form $\exists x_1 \ldots Q x_k\, A(x_1, \ldots, x_k)$ where the string of k number quantifiers at the front is alternating between existential and universal quantifiers and the matrix $A(x_1, \ldots, x_k)$ only contains bounded number quantifiers. Π^0_k is defined dually. Likewise, in the language of formal second order arithmetic, $\mathbf{Z}_2$, a formula $\exists X_1 \ldots Q X_k\, B(X_1, \ldots, X_k)$ is said to be Σ^1_k if the string of second order quantifiers alternates and $B(X_1, \ldots, X_k)$ does not contain any second order quantifiers (in the same vein one defines Π^1_k-formulas). Both languages will be viewed as part of the language of set theory by interpreting number quantifiers as ranging over the first limit ordinal ω, second order quantifiers as ranging over subsets of ω and interpreting addition and multiplication as ordinal addition and multiplication, respectively.

For theories T_1, T_2 and Γ a collection of formulas contained in $\mathcal{L}(T_1) \cap \mathcal{L}(T_2)$ (either directly or via the foregoing identification), define

$$T_1 \subseteq_\Gamma T_2 \quad :\Leftrightarrow \quad \forall \varphi \in \Gamma \, (T_1 \vdash \varphi \Rightarrow T_2 \vdash \varphi).$$

We say that T_2 *is* Γ*-conservative over* T_1 if $T_1 \subseteq_\Gamma T_2$ and $T_2 \subseteq_\Gamma T_1$. We convey this by writing $T_1 =_\Gamma T_2$.

A perhaps unexpected and non-intuitive case of theories related via $\trianglelefteq$ is the following.

Remark 2.3 Let $\mathsf{Con}(T)$ be the arithmetized predicate expressing (in a natural way) that the theory T is consistent. As a non-trivial example we have $\mathbf{PA} + \neg\mathsf{Con}(\mathbf{PA}) \ \trianglelefteq \ \mathbf{PA}$ (see [63]).

It is instructive to study the relation $\trianglelefteq$ on extensions T of $\mathbf{PA}$ that are primitive recursively axiomatized theories and have the same language as $\mathbf{PA}$. Such theories are *reflexive*, namely T proves the consistency of all its finite subtheories, i.e., $T \vdash \mathsf{Con}(T \restriction_k)$ for all $k \in \mathbb{N}$, where $T \restriction_k$ is the theory consisting of the first k-many axioms and $\mathsf{Con}(T \restriction_k)$ expresses its consistency.

The equivalence classes stemming from the relation of mutual relative interpretability, i.e., $S \equiv_{ri} T$, give rise to degrees of interpretability, yielding a dense distributive lattice (see [40]). The next result shows that for such theories $\trianglelefteq$ is closely related to the provability of consistency for finite fragments and the $\subseteq_{\Pi^0_1}$ relation.

Theorem 2.4 (Hilbert & Bernays, Feferman, Orey, Lindström, Guaspari)[3]
For extensions S, T of $\mathbf{PA}$ *as above, the following are equivalent.*

(i) $S \trianglelefteq T$.

(ii) For all $n \in \mathbb{N}$, $T \vdash \mathsf{Con}(S \restriction_n)$.

(iii) $S \subseteq_{\Pi^0_1} T$.

The previous theorem and also Remark 2.3 show that the notion of relative interpretability is of rather limited use when it comes to foundational theories, i.e. theories designed for the purpose of formalizing and capturing various parts of mathematics. Mutual reducibility should at least entail that both theories have the same algorithmic consequences, meaning that if one of them proves that an algorithm terminates on all integer inputs then the other should do so as well. In other words, at a minimum, mutual partial conservativity should hold for Π^0_2-statements. The latter class of statements contains many famous conjectures in mathematics (e.g. the twin prime conjecture). In mathematical logic one finds many interesting and surprising reductions between theories that turn out have the same Π^0_2-theorems. Moreover, there is also a sizable number of them that are not reducible to each other via relative interpretation. $S \trianglelefteq T$ requires that T accounts for S tout court. We owe it to Hilbert's genius that reductions need not be total. A theory T asserting the existence of 'ideal' objects, in the sense of his legendary program, may prove the same elementary

[3]For more details see [40, Section 6, Theorems 5 and 6]. Also the history of this theorem is related in [40]. A crucial tool in the proof is the arithmetization of Gödel's completeness theorem due to Hilbert and Bernays [34].

theorems (including Π^0_2-theorems) about elementary objects (such as the naturals) as a theory S without S being able to account for the 'ideal' objects of T.

Let us now return to the plan of providing a sample of interesting reducibility results for further discussions. From a technical point of view they were initially obtained by a plethora of methods from different branches of mathematical logic such as the constructible hierarchy, forcing, functional interpretation, recursively saturated models, realizability, topos theory, and cut elimination to name a few. Albeit all of these reductions will be expressed as partial conservativity results, they also turned out to be examples of pairs of theories that are proof-theoretically equivalent (a notion that will be introduced in 2.1), sometimes requiring completely new proofs.

Theorem 2.5 *(i) (Gödel 1938, 1940, [27, 28])* $\mathbf{ZFC} + \mathrm{GCH} =_{\mathcal{L}(\mathbf{PA})} \mathbf{ZF}$, *where* GCH *stands for the generalized continuum hypothesis.*

(ii) (Shoenfield 1961 [60]) $\mathbf{ZFC} + \mathrm{GCH} =_{\Pi^1_4} \mathbf{ZF}$.

(iii) (Platek [43] 1969, Kripke, Silver (both unpublished) 1969) $\mathbf{ZFC}+\mathrm{GCH} =_{\mathcal{L}(\mathbf{Z}_2)} \mathbf{ZFC}$.

(iv) (Parsons 1970 [42], Friedman 1976 [24]) $\mathbf{I}\Sigma^0_1 =_{\Pi^0_2} \mathbf{PRA}$ *and* $\mathbf{WKL}_0 =_{\Pi^0_2} \mathbf{PRA}$.[4]

(v) (Barwise & Schlipf 1975 [9]) $\Sigma^1_1\text{-}\mathbf{AC}_0 =_{\mathcal{L}(\mathbf{PA})} \mathbf{PA}$.

(vi) (Kripke, Solovay 1960s (independently), Felgner 1971 [23]) Von Neumann-Bernays-Gödel class set theory with the global axiom of choice, $\mathbf{NGBC}$*, is conservative over* $\mathbf{ZFC}$ *for formulas of* $\mathcal{L}(\mathbf{ZF})$.

(vii) (Kolmogorov 1925 [39], Gentzen 1933 [25], Gödel 1933 [26]) $\mathbf{PA}$ *is conservative over intuitionistic first order arithmetic (also known as Heyting arithmetic,* $\mathbf{HA}$*) for almost negative formulas.*[5]

(viii) (Kleene 1945 [38]) $\mathbf{HA}$ + *Church's thesis* $=_{\Pi^0_2} \mathbf{PA}$.

(ix) (Gödel 1958 [29]) $\mathbf{PA}$ *has a functional interpretation in Gödel's equational theory* $\mathbf{T}$ *of functionals of finite type. This is the known as the* Dialectica interpretation.

(x) (Goodman 1976, 1978 [30, 31]) $\mathbf{HA}^\omega + \mathbf{AC}_{type} =_{\mathcal{L}(\mathbf{HA})} \mathbf{HA}$. *Here* $\mathbf{HA}^\omega$ *denotes Heyting arithmetic in all finite types with* $\mathbf{AC}_{type}$ *standing for the collection of all higher type versions* $\mathbf{AC}_{\sigma\tau}$ *of the axiom of choice with* σ, τ *arbitrary finite types.*

(xi) (Barr 1974 [7]) Every classical geometric theory is conservative over its intuitionistic version with regard to geometric implications.[6]

(xii) (Rathjen 1993 [47], Setzer 1993, 1998 [58, 59]) $\Sigma^1_2\text{-}\mathbf{AC}+$ *Bar Induction*[7] *is reducible*

[4] Here $\mathbf{I}\Sigma^0_1$ stands for the fragment of $\mathbf{PA}$ with induction restricted to Σ^0_1-formulas, $\mathbf{PRA}$ for primitive recursive arithmetic, and $\mathbf{WKL}_0$ for a fragment of $\mathbf{Z}_2$ in which weak König's Lemma is the signature axiom. $\mathbf{WKL}_0$ is a very famous system in the the program of *reverse mathematics* (cf. [62]).

[5] A formula is almost negative if it does not contain $\vee$, and $\exists$ only immediately in front of Δ^0_0 formulas.

[6] A geometric theory (sometimes called coherent theory) is one whose axioms are geometric implications, i.e. universal closures of implications of the form $D_1 \to D_2$, where each D_i is a positive formula, i.e. a formula built up from atoms using solely conjunction, disjunction and existential quantification.

[7] Σ^1_2-$\mathbf{AC}$ is the schema $\forall x \exists Y\, B(x,Y) \to \exists Z \forall x\, B(x, Z_x)$ with $B(x,Y)$ of complexity Σ^1_2 and $Z_x = \{u \mid 2^x 3^u \in Z\}$. Bar induction stands for the schema of transfinite induction along a well-founded set relation for all second order formulas. This theory is very strong from the point of view of reverse mathematics. In reverse mathematics one aims to calibrate the strength of theorems from ordinary mathematics, using as a scale certain natural subsystems of $\mathbf{Z}_2$. It has turned out that there are five systems (dubbed the "big five") that occur most often. All of them are much weaker than Σ^1_2-$\mathbf{AC}$ + Bar Induction. Since the latter is reducible to Martin-Löf type theory, it follows that almost all Π^0_2-theorems of ordinary mathematics are constructively true (for more on this see [46]).

to Martin-Löf's 1984 type theory. More specifically, every Π^0_2-theorem of the former is a theorem of the latter system.

(xiii) *(Rathjen 1993 [47]) Let* **CZF** *be Constructive Zermelo-Fraenkel set theory,* REA *be the* regular extension axiom *(cf. [1, 2, 3, 4]), and* **KP** *be Kripke-Platek set theory (cf. [8]).*

$$\mathbf{CZF} \; =_{\Pi^0_2} \; \mathbf{KP}$$
$$\mathbf{CZF} + \mathrm{REA} \; =_{\Pi^0_2} \; \Sigma^1_2\text{-}\mathbf{AC} + \textit{Bar Induction.}$$

(xiv) *(Crosilla, Rathjen 2001 [17], Rathjen 2017 [45]) Let* $\mathbf{MLTT}^-$ *be intensional Martin-Löf type without W-types and* UA *be Voevodsky's univalence axiom. Let* $\mathbf{CZF}^-$ *be* **CZF** *without set induction and* INACC *be the axiom that every set is contained in an inaccessible set.*

$$\mathbf{MLTT}^- + \mathrm{UA} \; =_{\Pi^0_2} \; \mathbf{CZF}^- + \mathrm{INACC} =_{\Pi^0_2} \mathbf{ATR}_0$$

where $\mathbf{ATR}_0$ *stands for the theory of arithmetical transfinite recursion which is one of the central theories of reverse mathematics (see [62]).*

Of the foregoing cases of theory reduction, only the theories in (i), (ii), (iii), respectively stand in the relation $\equiv_{ri}$. Obviously, $\mathbf{ZFC}+\mathrm{GCH} \trianglelefteq \mathbf{ZF}$ holds by relativizing all quantifiers to L since **AC** and GCH are validated in the constructible hierarchy. Of course, the really interesting information contained in (ii) and (iii) is lost if rendered in this way: the point of (ii) is that if in principle we want to avoid AC[8] we can still use it and also the generalized continuum hypothesis if we want to prove a theorem expressible in the language of first order number theory and more generally if the theorem's statement does not have a complexity beyond Π^1_4 (which is actually pretty high). Likewise, if we are concerned with proving statements of second order arithmetic from **ZFC** we can freely use GCH as an extra hypothesis.

In all the other cases of theory reduction, i.e., (iv)-(xiv), the relation $\equiv_{ri}$, for various reasons, does not obtain between the respective theories:

1. The theory $\mathbf{WKL}_0$ of (iv) is finitely axiomatizable, and therefore, $\mathbf{WKL}_0 \trianglelefteq \mathbf{PRA}$ would yield an interpretation into a finite fragment of **PRA**, which is impossible.[9] Likewise in (v), Σ^1_1-$\mathbf{AC}_0$ is finitely axiomatizable whereas **PA** is not. The former theory has non-elementary speed-up over **PA**. The same considerations apply to **GBC** and **ZFC** in (vi).

2. Relative interpretability usually doesn't work across theories based on different logics. The double negation interpretation behind (vi) does not distribute over the logical connectives. The same holds for the realizability interpretation that underlies (viii). Also Gödel's functional interpretation of (ix) does not distribute over quantifiers. Similarly, the proof of (x) uses a combination of realizability and forcing, and there does not exist a plain interpretation of $\mathbf{HA}^\omega + \mathbf{AC}_{type}$ in **HA**.

3. Again in (xii)-(xiv), we have reductions across different logics which cannot be rendered relative interpretations. In (xiii) $\mathbf{CZF} + \mathrm{REA} \trianglelefteq \Sigma^1_2$-$\mathbf{AC}$ + Bar Induction is ruled out since the strength of $\mathbf{CZF} + \mathrm{REA}$ with classical logic is that of **ZF**. In the same vein, $\mathbf{CZF}^- + \mathrm{INACC} \trianglelefteq \mathbf{ATR}_0$ cannot hold as $\mathbf{CZF}^- + \mathrm{INACC}$ based on classical logic is stronger than **ZFC**.

[8] Recall that in 1905 several famous French analysts wanted to ban AC for uncountable families of sets from mathematics (see [37]).

[9] Indeed, $\mathbf{WKL}_0$ has non-elementary speed-up over **PRA**, cf. Definition 3.5 and [15].

4. The case of (xi) is quite different from the others in that it applies to a large collection of mathematical theories. Geometric theories are rather ubiquitous. They include all algebraic theories, such as group theory and ring theory, all essentially algebraic theories, such as category theory, the theory of fields, the theory of local rings, lattice theory, projective geometry and the theory of separably closed local rings (aka "strictly Henselian local rings"). The interest in geometric theories is not restricted to mathematics, e.g. it has been argued in [5] that Kant's transcendental logic is (infinite) geometric logic.

   (xi) was originally proved non-constructively, using topos-theoretic methods and the axiom of choice, but there is also an easy proof using Gentzen's Hauptsatz, i.e. cut elimination (cf. [18]). This also works for infinite geometric theories that are expressed in the language of infinitary logic and are defined in the same way as geometric theories except for allowing infinitary disjunctions in D_1 and D_2 (cf. [54]).

Remark 2.6 The examples (iv), (v), and (vi) from Theorem 2.5 are particularly pretty illustrations of Hilbert's "Methode der idealen Elemente" [33] and of what Husserl before him described as the "Durchgang durch das *Unmögliche*" and "Durchgang durch das *Imaginäre*"[10], respectively, in his 1901 Göttingen Doppelvortrag (see [16]). The idea here is that via an adjunction of ideal elements to the objects of a base theory S we arrive at an extended theory T with more objects. Crucially in these cases, putting ourselves in the place of S-people, we cannot speak about the objects of the T-people since there does not exists a translation by which quantification over the T-world can be construed as (a fancy) quantification over particular objects of our S-world.[11]

Having seen the failure of relative interpretability to account for the theory reductions in Theorem 2.5, we are desirous of finding one that has a greater scope and covers all of them.

2.1 Proof-theoretical reduction

The notion of proof-theoretical reduction first appeared in chapter 1 of [13] written by Feferman and Sieg. Subsequently it played a prominent role in Feferman's 1988 paper on relativized Hilbert programs.

All theories T considered in the following are assumed to contain a modicum of arithmetic. For definiteness let this mean that the system **PRA** of Primitive Recursive Arithmetic is contained in T, either directly or by translation.

Definition 2.7 Let T_1, T_2 be a pair of theories with languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, and let Φ be a (primitive recursive) collection of formulae common to both languages. Furthermore, Φ should contain the closed equations of the language of **PRA**.

We then say that T_1 is *proof-theoretically Φ-reducible to* T_2, written $T_1 \leq_\Phi T_2$, if there exists a primitive recursive function f such that

$$\mathbf{PRA} \vdash \forall\phi \in \Phi\, \forall x\, [\mathsf{Proof}_{T_1}(x,\phi) \;\rightarrow\; \mathsf{Proof}_{T_2}(f(x),\phi)]. \tag{1}$$

T_1 and T_2 are said to be *proof-theoretically Φ-equivalent*, written $T_1 \equiv_\Phi T_2$, if $T_1 \leq_\Phi T_2$ and $T_2 \leq_\Phi T_1$.

The appropriate class Φ is revealed in the process of reduction itself, so that in the statement of theorems we simply say that T_1 is *proof-theoretically reducible* to T_2 (written

[10] I.e., going through the *impossible* and the *imaginary*, respectively.

[11] For a partial realization of Hilbert's original program see Simpson [61].

$T_1 \leq T_2$) and T_1 and T_2 are *proof-theoretically equivalent* (written $T_1 \equiv T_2$), respectively. Alternatively, we shall say that T_1 and T_2 have the *same proof-theoretic strength* when $T_1 \equiv T_2$.

Remark 2.8 Since the reduction of theories is required to be provable in **PRA** and the latter is often considered to be co-extensional with finitism (see [65]), the notion of proof-theoretical reduction is synonymously used with *finistic reduction*. Alternatively, we will therefore say that T_1 is *finitistically reducible* to T_2 if T_1 is proof-theoretically reducible to T_2.

Remark 2.9 Feferman's notion of proof-theoretic reducibility in [22] is more relaxed in that he allows the reduction to be given by a T_2-recursive function f, i.e.

$$T_2 \vdash \forall \phi \in \Phi \, \forall x \, [\mathsf{Proof}_{T_1}(x, \phi) \ \rightarrow \ \mathsf{Proof}_{T_2}(f(x), \phi)]. \tag{2}$$

The disadvantage of (2) is that one forfeits the transitivity of the relation $\leq_\Phi$. In practice, however, proof-theoretic reductions always come with a primitive recursive reduction, so nothing seems to be lost by using the stronger notion of reducibility (for more details see [49]). Moreover, it has turned out that in actuality, i.e. in the cases studied in the literature, the class Φ always comprises the Π^0_2-sentences.

A paper [41] by Niebergall sets out to provide a good explication of the concept "the theory S is reducible to theory T". The results are summarized as follows: "*I show that one is quite naturally led to the well-known concept of "relative interpretability" as a candidate for a general explication of "reducibility". But what about alternative non-equivalent, maybe "incomparable", explications? Most of the rest of this paper is an attempt to show that there actually are no, and can be no, convincing* explicanta *of "reducibility" which are* very *different from relative interpretability.*" For sure, (relative) interpretability is a natural and venerable concept. But the conclusion that there can be no convincing explanation of "reducibility" very different from interpretability is surprising (also see [21]). Well, of course it depends on what "incomparable" and "very" in the quote mean. At any rate, many important phenomena, as we have seen in the discussion following Theorem 2.5, are not covered by this notion. More often than not, it turns out that it is either too general or too narrow. It's perhaps advisable to reflect on the wider purpose of these notions by looking at their natural habitat of application. Proof-theoretic reducibility is a crucial notion in the foundations of mathematics where the focus is on "natural" theories that serve as background frameworks for the formalization of chunks of mathematics. By contrast, the notion of interpretability is typically concerned with *all* theories of a certain kind, where reasonableness is not part of the concept of a theory. This leads to degree-theoretic studies of theories which focus on the algebraic structure of degrees of interpretability. In the latter enterprise the inclusion of "unreasonable" theories is of the essence (see [40]).

3 Weak Theories of Inductive Definitions

In this final chapter we come to the main part of this article which aims to explain how infinitary derivations can be used in proof theory without compromising finitistic reduction. The length of these derivations and the cost of their transformations is usually controlled by ordinals from an ordinal representation system. From the work of Gentzen in the 1930's up to the present time, this central theme is manifest in the assignment of 'proof theoretic ordinals' to theories, measuring their 'consistency strength' and 'computational power', and providing a scale against which those theories may be compared and classified. This branch of proof theory is known as *ordinal analysis* (see [56, 66, 49, 53] for more information).

Our goal will be achieved with the help of intuitionistic fixed point theories that are obtained from intuitionistic theories of inductive definitions by relaxing the least fixed point property associated with an inductively defined set.

A central feature of many constructive approaches to mathematics (e.g. Martin-Löf's type theory) is the insistence that sets of objects have to be generated according to rules. Formal theories of inductive definitions were proposed and investigated in the 1960s with the aim of assaying the scope of constructive foundations.[12]

From a set-theoretic point of view, any monotone operator on $\mathbb{N}$ gives rise to an inductive definition. A monotone operator is a map Γ that sends a set $X \subseteq \mathbb{N}$ to a subset $\Gamma(X)$ of $\mathbb{N}$ and is monotone, i.e. $X \subseteq Y \subseteq \mathbb{N}$ implies $\Gamma(X) \subseteq \Gamma(Y)$. Owing to monotonicity, the operator Γ will have a least fixed point $I_\Gamma \subseteq \mathbb{N}$, i.e. $\Gamma(I_\Gamma) = I_\Gamma$ and for every other fixed point X of Γ (i.e. $\Gamma(X) = X$) one has $I_\Gamma \subseteq X$. Set-theoretically I_Γ is obtained by iterating Γ through the ordinals,

$$\Gamma^0 = \emptyset, \quad \Gamma^1 = \Gamma(\Gamma^0), \quad \Gamma^\alpha = \Gamma(\bigcup_{\xi<\alpha} \Gamma^\xi).$$

The axioms of set theory guarantee that one finds an ordinal τ such that $\Gamma(\Gamma^\tau) = \Gamma^\tau$, and the set Γ^τ will be the least fixed point, on account of Γ's monotonicity.

If one adds a new 1-place predicate symbol P to the language of arithmetic, one can describe the so-called positive arithmetical operators. They are of the form

$$\Gamma_A(X) = \{n \in \mathbb{N} \mid A(n, X)\}$$

where $A(x, P)$ is a formula of the language of **PA** upon augmentation by P in which the predicate P occurs only positively. The syntactic condition of positivity then ensures that the operator Γ_A is monotone. Below we shall focus on special positive arithmetic operators. More generally, we shall allow for iterated inductive definitions.

Definition 3.1 The language $\mathcal{L}_1(Q)$ extends that of Heyting Arithmetic, by means of a new unary predicate symbol Q, whereas $\mathcal{L}_1(Q, P)$ denotes the extension by two unary predicate symbols P and Q. For convenience we assume that the free variables $v_0, v_1, v_2, \ldots$ and the bound variables $x_0, x_1, x_2, \ldots$ are syntactically different. Terms are built from free variables, constants and function symbols in the usual way.

The set of *strictly positive operator forms* is characterized by the strictly positive formulas of $\mathcal{L}_1(Q, P)$. The latter are generated by the following clauses:

1. Every formula of $\mathcal{L}_1(Q)$ is a strictly positive formula.
2. For every term t the formula $P(t)$ is a strictly positive formula.
3. The strictly positive formulas are closed under $\exists, \forall, \land$ and $\lor$.
4. If A is an $\mathcal{L}_1(Q)$-formula and B is a strictly positive formula, then $A \to B$ is a strictly positive formula.

[12] At the time a particular focus of proof theorists was Spector's 1960 functional interpretation of $\mathbf{Z_2}$ via bar recursive functionals. The question whether such functionals are acceptable on constructive grounds was of great interest to proof theorists, constructivists and logicians who wanted to plumb the boundaries of constructivism (e.g. Gödel).

A *strictly positive operator form* is a strictly positive formula that contains at most v_0 and v_1 as free variables. The set of *accessibility operator forms* is the subset of the strictly positive operator forms that have the form

$$A \wedge \forall z[B(z) \to P(z)]$$

with $A, B \in \mathcal{L}_1(Q)$. The set of *strongly positive operators forms* is the set of strictly positive operators forms that are generated by the clauses 1–3 (omitting 4).

Definition 3.2 The language $\mathcal{L}_{strict}$ arises from $\mathcal{L}(\mathbf{HA})$ by adding a new unary predicate P^A for every strictly positive operator form $A(P, Q, v_0, v_1)$ of $\mathcal{L}_1(Q, P)$.

To formulate axioms pertaining to P^A we use the following shorthands: $P^A_s(t)$ stands for $P^A(\langle t, s\rangle)$ and $P^A_{<s}(t)$ for $t = \langle (t)_0, (t)_1\rangle \wedge (t)_1 < s \wedge P^A(t)$. Here $\langle \cdot, \cdot \rangle$ denotes a primitive recursive pairing function with projections $(\cdot)_0$ and $(\cdot)_1$.

The theory $\widehat{\mathbf{ID}}^i_n(strict)$ has $\mathcal{L}_{strict}$ as its language and is based on intuitionistic logic. It comprises all axioms of $\mathbf{HA}$ with the induction scheme extended to all $\mathcal{L}_{strict}$-formulas and has for every strictly positive operator form $A(P, Q, v_0, v_1)$ the fixed point axiom:

$$(\forall y < n)(\forall x)[P^A_y(x) \leftrightarrow A(P^A_y, P^A_{<y}, x, y)].$$

Let $\widehat{\mathbf{ID}}^i_{<\omega}(strict) := \bigcup_{n\in\mathbb{N}} \widehat{\mathbf{ID}}^i_n(strict)$.
The theories $\widehat{\mathbf{ID}}^i_n(acc)$, $\widehat{\mathbf{ID}}^i_{<\omega}(acc)$, $\widehat{\mathbf{ID}}^i_n(strong)$, and $\widehat{\mathbf{ID}}^i_{<\omega}(strong)$ are defined analogously by restricting the fixed point axioms to the pertaining operator forms.

Remark 3.3 If one adds the schema

$$(P^A\text{-Induction}) \qquad (\forall x)[A(\varphi, x) \to \varphi(x)] \to (\forall x)[P^A(x) \to \varphi(x)],$$

for all formulas $\varphi(v)$, to the axioms of $\widehat{\mathbf{ID}}^i_1(strict)$ one arrives at a rather strong theory, denoted by $\mathbf{ID}^i_1(strict)$, which is of the same strength as its classical version $\mathbf{ID}_1(strict)$. P^A-Induction schematically express the leastness of the fixed point predicate P^A.

The classical theories $\widehat{\mathbf{ID}}_n(strict)$ are much weaker than $\mathbf{ID}^i_1(strict)$. Yet, $\widehat{\mathbf{ID}}_1(strict)$ is stronger than $\mathbf{PA}$ and $\widehat{\mathbf{ID}}_{n+1}(strict)$ is stronger than $\widehat{\mathbf{ID}}_n(strict)$. That the intuitionistic theories exhibit a completely different behavior, was first observed by Wilfried Buchholz (see [10]). Indeed, it turned out that all the intuitionistic fixed points hierarchies collapse to $\mathbf{HA}$.

Buchholz' work in [10] spawned several extension results.

Theorem 3.4 *(i) (Buchholz [10])* $\mathbf{HA} =_{\Pi^0_2} \widehat{\mathbf{ID}}^i_n(strong)$.

(ii) (Arai [6]) $\widehat{\mathbf{ID}}^i_{<\omega}(strong)$ *is conservative over* $\mathbf{HA}$

(iii) (Rüede, Strahm [64]) The theories $\widehat{\mathbf{ID}}^i_1(strict)$ *and* $\widehat{\mathbf{ID}}^i_{<\omega}(acc)$ *are conservative over* $\mathbf{HA}$ *for all negative and* Π^0_2*-sentences of* $\mathcal{L}(\mathbf{HA})$.

(iv) The above results still obtain when one adds schemata $\mathrm{TI}(\prec)$ *for primitive recursive* $\prec$ *to the above theories.*

(v) All of the theories of (i), (ii), (iii) are finistically reducible to $\mathbf{HA}$.

(vi) All of the theories of (i), (ii), (iii), when augmented by $\mathrm{TI}(\prec)$*, are finistically reducible to* $\mathbf{HA} + \mathrm{TI}(\prec)$.

Proof: Buchholz' proof uses an interpretation of $\widehat{\mathbf{ID}}^i_n(strong)$ in $\mathbf{HA}$ + *Church's thesis* which is known to be conservative over $\mathbf{HA}$ for Π^0_2-formulas.

(ii) Arai's extension uses the techniques that were used in Goodman's theorem to show that $\mathbf{HA}^\omega$ + AC, i.e., Heyting arithmetic in all finite types together with the axiom of choice for all type levels, is conservative over $\mathbf{HA}$.

(iii) Rüede and Strahm's proof first employs recursive realizability to reduce $\widehat{\mathbf{ID}}^i_n(strict)$ to $\widehat{\mathbf{ID}}^i_n(acc)$ (following Buchholz [12]) and subsequently models $\widehat{\mathbf{ID}}^i_n(acc)$ in a classical theory with arithmetic comprehension, using an argument from [20], where it is shown that a Π^0_2 fixed point can be obtained for the accessibility inductive definition of Kleene's $\mathcal{O}$.

(iv) follows from the fact that all transformations between the pertaining theories also work in the presence of TI($\prec$).

(v) and (vi). The methods employed are direct proof transformations that give rise to primitive recursive functions. More details will be provided in Theorem 3.9. □

3.1 Measuring the complexity of reductions

The foregoing reductions can be engineered by primitive recursive functions. Still there remains the question whether these results can be obtained by stricter notions of reduction. This has been investigated by Michael Toppel in his thesis [67]. To discuss the cost of proof transformations, we recall some notions pertaining to length of proofs and proof-growth. Taking any of the familiar proof systems (e.g. Hilbert-style, natural deduction, sequent calculus, Schütte calculus), the length of a proof (aka derivation and deduction), $\mathcal{D}$, will be denoted by $|\mathcal{D}|$. The latter is the total length of $\mathcal{D}$, i.e., counting the total number of all occurrences of symbols in $\mathcal{D}$.

A function $g : \mathbb{N} \to \mathbb{N}$ has *polynomial growth rate* if it is eventually dominated by a polynomial $p(x)$ in x with coefficients in $\mathbb{N}$, i.e., there exists $m \in \mathbb{N}$ such that $g(x) \leq p(x)$ for all $x > m$.

By 2^n_k we denote the tower of k many two's with the number n at the top, i.e., $2^n_0 = n$ and $2^n_{r+1} = 2^{2^n_r}$. A function $g : \mathbb{N} \to \mathbb{N}$ is said to have (at most) *elementary growth* (or *Kalmar elementary growth*) if there exists a number m such that g is eventually dominated by the function $x \mapsto 2^x_m$. By contrast, we say that g has *super-exponential growth rate* if there exists a polynomial $p(x)$ in x with coefficients in $\mathbb{N}$ such that $g(x) \leq p(2^x_x)$ for all sufficiently large $x \in \mathbb{N}$, but g does not have elementary growth.

Definition 3.5 If S is a theory and $S \vdash \psi$ let $\mathcal{D}^\psi_S$ denote the shortest proof of ψ in S.

Let S and T be theories such that S is a subtheory of T. Let Γ be a set of formulas of $\mathcal{L}(S)$.

(i) We say that T has at most *polynomial speed-up* over S with respect to Γ if there exists a polynomial $p(x)$ with coefficients in $\mathbb{N}$ such that

$$|\mathcal{D}^\psi_S| < p(|\mathcal{D}^\psi_T|)$$

holds for all $\psi \in \Gamma$ such that $S \vdash \psi$.

(ii) T is said to have *super-exponential speed-up* over the theory S with regard to Γ if there exists a sequence ψ_n of Γ-theorems of S such that there is no function g with elementary growth rate satisfying $|\mathcal{D}^{\psi_n}_S| < g(|\mathcal{D}^{\psi_n}_T|)$ for all n, however, there exists a function f with super-exponential growth rate such that for all n, $|\mathcal{D}^{\psi_n}_S| < f(|\mathcal{D}^{\psi_n}_T|)$.

To state the results of Theorem 3.4 in a precise quantitative form, it is important to pay attention to the way how function symbols for primitive recursive functions are handled in the language.

Definition 3.6 The set of primitive recursive function symbols (p.r.f.s) is inductively defined as follows:

1. The symbol for the successor function S is a symbol for a p.r.f.s of arity 1.
2. The symbol for the constant-0-function of arity n, C^n, is a p.r.f.s of arity n.
3. The symbol for the projection to the k-th of n inputs, P^n_k, is a p.r.f.s of arity n.
4. If $f_1^{n_1}, \ldots, f_{n_1+1}^{n_{n_1+1}}$ are p.r.f.s with their arities shown, then $[Sub(f_1^{n_1}, \ldots, f_{n_1+1}^{n_{n_1+1}})]^n$, where $n := \max\{n_2, \ldots, n_{n_1+1}\}$, is a p.r.f.s of arity n.[13]
5. If $f_1^{n_1}, f_2^{n_2}$ are p.r.f.s with their arities shown, then $[Rec(f_1^{n_1}, f_2^{n_2})]^n$, where $n := \max\{n_1 + 1, n_2\}$, is a p.r.f.s of arity n.

Definition 3.7 The axioms of the p.r.f.s are the universal closures of the following formulas.

1. For S:
$$\neg S(x) = \bar{0}$$
$$S(x) = S(y) \rightarrow x = y$$
2. For a C^n:
$$C^n(x_1, \ldots, x_n) = \bar{0}$$
3. For a P^n_k:
$$P^n_k(x_1, \ldots, x_n) = x_k$$
4. For a $f^n \equiv [Sub(f_1^{n_1}, \ldots, f_{n_1+1}^{n_{n_1+1}})]^n$:
$$f^n(x_1, \ldots, x_n) = f_1^{n_1}(f_2^{n_2}(x_1, \ldots, x_{n_2}), \ldots, f_{n_1+1}^{n_{n_1+1}}(x_1, \ldots, x_{n_{n_1+1}}))$$
5. For a $f^n \equiv [Rec(f_1^{n_1}, f_2^{n_2})]^n$:
$$f^n(0, x_1, \ldots, x_{n-1}) = f_1^{n_1}(x_1, \ldots, x_{n_1})$$
$$f^n(S(y), x_1, \ldots, x_{n-1}) = f_2^{n_2}(f^n(y, x_2, \ldots, x_{n-1}), y, x_3, \ldots, x_{n_2}).$$

We require additional measures of complexity for terms.

Definition 3.8 Assume f is a p.r.f.s. The degree of f, $\mathrm{dg}(f)$, is defined as follows:

1. If f is S, C^n or P^n_k, then $\mathrm{dg}(f) = 1$.
2. If $f \equiv [Sub(f_1^{n_1}, \ldots, f_{n_1+1}^{n_{n_1+1}})]^n$, then
$$\mathrm{dg}(f) = \sum_{i=0}^{n_1+1} \mathrm{dg}(f_i) + 1.$$

[13] Note that there have to be n_1 function symbols substituted into a function symbol with arity n_1.

3. If $f \equiv [Rec(f_1^{n_1}, f_2^{n_2})]^n$, then

$$\mathrm{dg}(f) = \mathrm{dg}(f_1) + \mathrm{dg}(f_2) + 1.$$

For a term t, define $\mathrm{dg}(t)$ to be the largest degree of a p.r.f.s occurring in t.
The rank of t, $\mathrm{rk}(t)$, is defined as follows.

1. If $t \equiv x$ or $t \equiv \bar{0}$, then $\mathrm{rk}(t) = 0$.
2. If $t \equiv f(t_1, \ldots, t_n)$, then $\mathrm{rk}(t) = n \cdot (\max\{\mathrm{rk}(t_i) | 1 \leq i \leq n\} + 1)$.

For naturals k, n, m define $k_0^m := m$ and $k_{n+1}^m := k^{k_n^m}$.

Theorem 3.9 *Assume that both $\widehat{\mathbf{ID}}_n^i(strict)$ and $\mathbf{HA}$ are formulated with the same primitive recursive function symbols of degree $\leq g$ (thus finitely many). Moreover, let $\mathcal{W}$ be a family of binary Σ_0^0 definable relations of the language of $\mathbf{HA}$. By $\mathrm{TI}(\prec)$ we denote the schema of transfinite induction along $\prec$ for all formulas of $\widehat{\mathbf{ID}}_n^i$ and $\mathrm{TI}_{arith}(\prec)$ denotes the schema of transfinite induction along $\prec$ for all formulas of $\mathbf{HA}$. Then there exists a polynomial $P(X)$ with positive integer coefficients such that for all almost negative and Π_2^0 formulas A of $\mathbf{HA}$,*

$$\widehat{\mathbf{ID}}_n^i(strict) + \{\mathrm{TI}(\prec) \mid \prec \in \mathcal{W}\} \overset{m}{\vdash} A \ \Rightarrow\ \mathbf{HA} + \{\mathrm{TI}_{arith}(\prec) \mid \prec \in \mathcal{W}\} \overset{P(m_{2g}^{2g})}{\vdash} A\,.$$

Proof: This a consequence of the following results from [67]: Lemma 4.1.6, Theorem 4.1.12, Theorem 4.2.8 and Theorem 4.3.1. □

The foregoing Theorem shows that the cost of proof transformation between $\widehat{\mathbf{ID}}_\omega^i(strict)$ and $\mathbf{HA}$ is controlled by an elementary function. As a special case, note that if we restrict the vocabulary to the symbols for the functions $0, S, +, \cdot$, then their degree is ≤ 3 and thus $x \mapsto P(x_6^6)$ is an upper bound for the cost of this proof transformation. □

The foregoing theorem has important consequences as far as ordinal analysis is concerned in that it can be used to show that proof-theoretical reductions between theories established via ordinal analysis are even strictly finitistic. Here strictly finitistic means that the proof transformations are controlled by elementary functions and can formally be proved in elementary recursive arithmetic. The next subsection will provide some details as to how this is achieved.

3.2 $\widehat{\mathbf{ID}}_{<\omega}^i(strict)$ plus $\mathrm{TI}(\prec)$ as a metatheory for ordinal analysis

That the theory in the title of this subsection provides a background theory in which the proof theory of infinitary derivations with ordinal labels from an ordinal representation system (with ordering $\prec$) can be carried out in an almost unencumbered style, was first pointed out by Wilfried Buchholz in [10].

Hitherto the strategy for arguing that infinitary proof theory can be carried out in an arithmetic theory beefed up by transfinite induction was based on the observation that it often suffices to focus on recursive proof–trees instead of arbitrary derivations. In many cases it is of course possible to restrict oneself to codes for recursive proof trees (see e.g. [57, 44]). One of the drawbacks of this approach is that it requires a lot of encoding of metamathematical notions and is cumbersome to carry out in detail.[14] As a result, one

[14]This is also the reason why complete proofs of Gödel's second incompleteness theorem are hard to find in the literature.

frequently resorts to hand waving rendering it difficult to get precise bounds for the length of proof transformations. The problem gets aggravated when the proof system has even more complex infinitary proof rules than the ω-rule. The ordinal analysis for Kruskal's theorem by Rathjen and Weiermann in [55] used derivations with Buchholz' Ω-rule, the formalization of which usually requires an iterated inductive definition, making it far from clear that one can resort to recursive derivations in this case. A treatment of the latter in $\mathbf{ACA}_0$ was sketched at the end of [55], but the paper's two finishing lines indicate that more needs to be done.

> *To carry out all the details of this constructivization would mean to produce another lengthy paper. But it is high time that we finished this paper; so we simply quit at this point.*[55, pp. 87–88]

With the help of $\widehat{\mathbf{ID}}^i_2(strict)$, a complete formalization could finally be furnished by M. Toppel in [67, chap. 4].

We will only indicate the general procedure for the Schütte-style (cf. [56]) ordinal analysis of a theory using a derivability predicate $\mathcal{D}(\alpha, \rho, \Gamma)$ signifying that the sequent Γ is derivable with length α and *cut-rank* ρ, where the "ordinals" come from some representation system ORT with ordering $\prec$. Such a predicate is usually defined by transfinite recursion on α according to the following pattern:

$$
\begin{array}{lll}
(*) \quad \mathcal{D}(\alpha, \rho, \Gamma) & \Leftrightarrow & \alpha \in \mathrm{ORT} \text{ and } \Gamma \text{ is either an axiom or the conclusion} \\
& & \text{of an inference with premises } (\Gamma_\iota)_{\iota \in I} \text{ such that} \\
& & \text{for every } \iota \in I \text{ there exists } \alpha_i \prec \alpha \text{ satisfying } \mathcal{D}(\alpha_\iota, \rho, \Gamma_\iota), \\
& & \text{and if the inference is a cut the cut-formulas have rank } \prec \rho.
\end{array}
$$

Alternatively, one can view $(*)$ as a fixed point axiom in the theory $\widehat{\mathbf{ID}}^i_1$. The "well-foundedness" of the derivability predicate $\mathcal{D}$ is then guaranteed in the background theory $\widehat{\mathbf{ID}}^i_1 + \mathrm{TI}(\prec)$ by the fact that the premises always have smaller ordinal "tags" than the conclusion, i.e., all the relevant results about this notion of derivability can be proved in the customary way by transfinite induction on the ordinal length and cut-rank.

Let's discuss a typical example of an ordinal analysis of a classical theory T. The starting point is an ordinal representation system that gives rise to a primitive recursive (actually elementary) well-ordering $<$ together with a canonical sequence of initial segments $<_k$ of $<$ such that $<= \bigcup_{k \in \mathbb{N}} <_k$. From a proof $\mathcal{D}$ of a Π^0_2-statement A in T one effectively determines an initial segment $<_k$ of $<$ such that there is a cut free proof of A of length $<_k$ in an infinitary proof system. It is a fact that for many ordinal analyses

$$\widehat{\mathbf{ID}}^i_n(strict) + \{\mathrm{TI}(<_k) \mid k \in \mathbb{N}\}$$

provides an adequate metatheory for handling infinitary proofs (or rather provability with ordinal bounds) and the cut elimination procedure (where actually $n = 1$ or $n = 2$). In other words, if a background theory is expressive enough to be able to formalize infinite derivability and comes equipped with sufficient amounts of transfinite induction, then intuitionistic logic suffices to carry out an ordinal analysis of T. As a result, $\widehat{\mathbf{ID}}^i_n(strict) + \{\mathrm{TI}(<_k) \mid k \in \mathbb{N}\}$ also proves A and with the help of Theorem 3.9 it follows that $\mathbf{HA} + \{\mathrm{TI}(<_k) \mid k \in \mathbb{N}\}$ proves A. Thus T is proof-theoretically (or finitistically) reducible to $\mathbf{HA} + \{\mathrm{TI}(<_k) \mid k \in \mathbb{N}\}$.

Ordinal analyses, such as the the ones in Schütte's book [56] (especially those in chapter VIII. Predicative Analysis), can be easily seen to just require a metatheory of the form

$\widehat{\mathbf{ID}}^{i}_{<\omega}(strict)$ plus TI($\prec$). Also ordinal analyses of some of the strongest theories (e.g. [11, 48, 51, 52]) just require an intuitionistic metatheory.

Whereas most ordinal analysis in the literature can be seen to just require an intuitionistic background theory (sometimes after removing some unnecessary classical arguments), there are exceptions to this pattern. For instance in the book by Buchholz and Schütte [14], the treatment becomes quite enmeshed with classical logic at the meta level. It is based on the Ω_n-rules (and even Ω_α-rules) which very likely could be handled in the theories $\widehat{\mathbf{ID}}^{i}_{n+1}(strict)$ and $\widehat{\mathbf{ID}}^{i}_{\alpha+1}(strict)$, respectively, but the authors also frequently use the principle

$$\exists\beta\, P(\beta) \to \exists\beta\, [P(\beta) \wedge \forall\gamma < \beta\, \neg P(\gamma)]$$

which is only classically equivalent to transfinite induction. However, there are other ordinal analyses of the theories treated in [14] that avoid this problem, e.g. the ones from [11] and [48].

Acknowledgements A substantial part of this paper was written while the first author participated in the trimester *Types, Sets and Constructions* May - August 2018 at the Hausdorff Research Institute for Mathematics (HIM), University of Bonn. His visit was supported by the HIM. Both, the support and the hospitality of HIM, are gratefully acknowledged.

This publication was made possible through the support of a grant from the John Templeton Foundation ("A new dawn of intuitionism: mathematical and philosophical advances," ID 60842).[15]

The second author was supported by a University Research Scholarship of the University of Leeds.

References

[1] P. Aczel: *The type theoretic interpretation of constructive set theory: Choice principles.* In: A.S. Troelstra and D. van Dalen, editors, *The L.E.J. Brouwer Centenary Symposium*, North Holland, Amsterdam (1982) 1–40.

[2] P. Aczel: *The type theoretic interpretation of constructive set theory: Inductive definitions.* In: R.B. Marcus et al. (eds.): *Logic, Methodology and Philosophy of Science VII* (North Holland, Amsterdam, 1986) 17–49.

[3] P. Aczel, M. Rathjen: *Notes on constructive set theory*, Technical Report 40, Institut Mittag-Leffler (The Royal Swedish Academy of Sciences, 2001). `http://www.mittag-leffler.se/preprints/0001/`, Preprint No. 40.

[4] P. Aczel, M. Rathjen: *Notes on constructive set theory*, Preprint (2010) 243 pages. http://www1.maths.leeds.ac.uk/ rathjen/book.pdf

[5] T. Achourioti, M. van Lambalgen: *Kant's Logic Revisited*. IfCoLog Journal of Logics and their Applications 4 (2017) 845–865.

[6] T. Arai: *Some results on cut-elimination, provable well-orderings, induction and reflection.* Annals of Pure and Applied Logic 95 (1998) 93–184.

[7] M. Barr: *Toposes without points.* J. Pure and Applied Algebra 5 (1974) 265–280.

[8] J. Barwise: *Admissible Sets and Structures* (Springer, Berlin 1975).

[9] J. Barwise, J. Schlipf: *On recursively saturated models of arithmetic.* Lecture Notes in Mathematics 498 (Springer, Berlin, 1975) 42–55.

[15] The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

[10] W. Buchholz: *An intuitionistic fixed point theory.* Archive for Mathematical Logic 37 (1997) 21–27.

[11] W. Buchholz: *A simplified version of local predicativity.* in: P. Aczel, H. Simmons, S. Wainer (eds.), *Leeds Proof Theory 90* (Cambridge University Press, Cambridge, 1993) 115-147.

[12] W. Buchholz: *The* $\Omega_{\mu+1}$*-Rule.* In: W. Buchholz, S. Feferman, W. Pohlers, W. Sieg, *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, Springer-Verlag, Berlin (1981).

[13] W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981).

[14] W. Buchholz and K. Schütte: *Proof Theory of Impredicative Subsystems of analysis.* (Bibliopolis, Napoli, 1988).

[15] P. Caldon, A. Ignjatović: *On Mathematical Instrumentalism.* In: Journal of Symbolic Logic, Vol. 70 (2005) pp. 778-794.

[16] S. Centrone: *Husserls Doppelvortrag in der Mathematischen Gesellschaft in Göttingen 1901.* In: C. Beyer & K. Cramer (Hg.), Edmund Husserl 1859-2009. Beiträge aus Anlass der 150. Wiederkehr des Geburtstages des Philosophen, Abhandlungsreihe der Akademie der Wissenschaften zu Göttingen ADW 14 (De Gruyter, Berlin 2011) 107–128.

[17] L. Crosilla, M. Rathjen: *Inaccessible set axioms may have little consistency strength.* Annals of Pure and Applied Logic 115 (2002) 33–70.

[18] R. Dyckhoff, S. Negri: *Geometrization of first-order logic.* Bull. Symbolic Logic 21 (2015) 123–163.

[19] S. Feferman: *How is it that finitary proof theory became infinitary?* Abstract for the ASL Logic Colloquium 1994 in Claremont-Ferrand.

[20] S. Feferman: *Iterated inductive fixed-points theories: application to Hancock's conjecture.* In *Patras Symposium*, G. Metakides (ed.) (North-Holland, Amsterdam, 1982) 171–196.

[21] S. Feferman: *Does Reductive Proof Theory have a viable Rationale?.* In: Erkenntnis, Vol. 53 (2000) pp. 63-96

[22] S. Feferman: *Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions.* In: Journal of Symbolic Logic, Vol. 53 (1988) pp. 364-384.

[23] U. Felgner: *Comparison of the axioms of local and universal choice.* Fundamenta Mathematicae 71 (1971) 43–62.

[24] H. Friedman: *Systems of second order arithmetic with restricted induction , I, II (abstracts).* J. of Symbolic Logic 41 (1976) 557–559.

[25] G. Gentzen: *Über das Verhältnis zwischen intuitionistischer und klassischer Logik* (1933). Originally to appear in the Mathematische Annalen, reached the stage of galley proofs but was withdrawn. It was finally published in Arch. Math. Logik 16 (1974) 119–132.

[26] K. Gödel: *Zur intuitionistischen Arithmetik und Zahlentheorie.* Ergebnisse eines mathematischen Kolloquiums 4 (1933) 34-38.

[27] K. Gödel: *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis.* Proceedings of the National Academy of Sciences of the United States of America. 24 (1938) 556–557.

[28] K. Gödel: *The Consistency of the Continuum-Hypothesis.* (Princeton University Press, Princeton, New Jersey, 1940).

[29] K. Gödel: *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes.* Dialectica 12 (1958) 280–287.

[30] N. Goodman: *The theory of the Gödel functionals.* J. Symbolic Logic 41 (1976), 574–583.

[31] N. Goodman: *Relativized Realizability in Intuitionistic Arithmetic of All Finite Types*, J. of Symbolic Logic 43 (1978).

[32] J. van Heijenoort (ed.): *From Frege to Gödel. A Source Book in Mathematical Logic 1879-1931.* (Harvard university Press, Cambridge Mass., 1967).

[33] D. Hilbert: *Über das Unendliche.* Mathematische Annalen 95 (1926) 161-190. English translation in: J. van Heijenoort (ed.): *From Frege to Gödel. A Source Book in mathematical logic, 1897-1931.*(Harvard University Press, Cambridge, Mass.,1967).

[34] D. Hilbert and P. Bernays: *Grundlagen der Mathematik II* (Springer, Berlin, 1939).

[35] P. Hinman: *Recursion-theoretic hierarchies* (Springer, Berlin, 1978).

[36] E. Husserl: *Studien zur Arithmetik und Geometrie.* Texte aus dem Nachlass 1886–1901. Husserliana Band XXI, I. Strohmeyer (Hrsg.), (Martinus Nijhoff, Den Haag, 1983).

[37] H. Jervell: *From the axiom of choice to choice sequences*, Nordic Journal of Philosophical Logic 1 (1996) 95-98.

[38] S.C. Kleene: *On the interpretation of intuitionistic number theory.* Journal of Symbolic Logic 10 (1945) 109–124.

[39] A.N. Kolmogorov: *On the principle of the excluded middle (Russian).* Mat.Sb. 32 (1925) 646–667. Translated in [32, 414–437].

[40] P. Lindström: *Aspects of Incompleteness.* Lecture Notes in Logic 10, Springer-Verlag, Berlin (1997)

[41] K.-G. Niebergall: *On the logic of reducibility: Axioms and Examples.* Erkenntnis 53 (2000) 27–61.

[42] C. Parsons: *On a number-theoretic choice schema and its relation to induction.* In: J. Myhill, A. Kino, R.E. Vesley (eds.), Intuitionism and Proof Theory (North-Holland, Amsterdam, 1970) 459–473.

[43] R.A. Platek: *Eliminating the Continuum Hypothesis.* J. of Symbolic Logic 34 (1969) 219–225.

[44] W. Pohlers: *Proof–theoretical analysis of ID_ν by the method of local predicativity*, in: W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981) 261–357.

[45] M. Rathjen: *Proof Theory of Constructive Systems: Inductive Types and Univalence.* In: Feferman on Foundations - Logic, Mathematics, Philosophy. Ed. by G. Jäger and W. Sieg. Vol. 13. Outstanding Contributions to Logic. (Springer, Cham, 2017) 385–419.

[46] M. Rathjen: *The constructive Hilbert programme and the limits of Martin-Löf type theory.* Synthese 147 (2005) 81–120.

[47] M. Rathjen: *The strength of some Martin–Löf type theories.* Preprint, Department of Mathematics, Ohio State University (1993) 39 pages.

[48] M. Rathjen: *Proof theory of reflection.* Annals of Pure and Applied Logic 68 (1994) 181–224.

[49] M. Rathjen: *The Realm of Ordinal Analysis.* In: Set and Proofs, London Math. Soc. Lecture Note Ser. 258, edited by S. B. Cooper and J. K. Truss, Cambridge University Press, Cambridge (1999), 219–279.

[50] M. Rathjen: *The constructive Hilbert programme and the limits of Martin-Löf type theory* Synthese 147 (2005) 81 - 120.

[51] M. Rathjen: *An ordinal analysis of stability*, Archive for Mathematical Logic 44 (2005) 1–62.

[52] M. Rathjen: *An ordinal analysis of parameter-free Π^1_2 comprehension.* Archive for Mathematical Logic 44 (2005) 263–362.

[53] M. Rathjen: *The Art of Ordinal Analysis.* In: M. Sanz-Solé, J. Soria, J.L. Varona, J. Verdera (eds.): *Proceedings of the International Congress of Mathematicians Madrid 2006*, Volume II (European Mathematical Society, 2006) 45–69.

[54] M. Rathjen: *Remarks on Barr's theorem. Proofs in Geometric Theories.*. In: Concepts of Proof in Mathematics, Philosophy, and Computer Science. Ed. by D. Probst and P. Schuster. Vol. 6. Ontos Mathematical Logic (Walter de Gruyter, Berlin, 2016) 347–374.

[55] M. Rathjen and A. Weiermann: *Proof–theoretic investigations on Kruskal's theorem,* Annals of Pure and Applied Logic 60 (1993) 49–88.

[56] K. Schütte: *Proof Theory.* (Springer, Berlin, 1977).

[57] H. Schwichtenberg: *Proof theory: Some applications of cut-elimination.* In: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam, 1977) 867–895.

[58] T. Setzer: *Proof theoretical strength of Martin–Löf type theory with W–type and one universe* (Thesis, University of Munich, 1993).

[59] A. Setzer: *A well-ordering proof for the proof theoretical strength of Martin-Löf type theory*, Annals of Pure and Applied Logic 92 (1998) 113–159.

[60] J.R. Shoenfield: *The problem of predicativity.* In: essays on the foundation of mathematics (Amsterdam, North-Holland, 1961) 132–139.

[61] S. Simpson: *Partial realizations of Hilbert's program.* Journal of Symbolic Logic 53 (1988) 349–363.

[62] S. Simpson: *Subsystems of second order arithmetic*, second edition, (Cambridge University Press, New York, 2009).

[63] C. Smorynski: *The incompleteness theorems.* In: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam, 1977) 821–865.

[64] C. Rüede, T. Strahm: *Intuitionistic Fixed Point Theories for Strictly Positive Operators.* In: Mathematical Logic Quarterly, Vol. 48, Issue 2 (2002), pp. 195-202.

[65] W. Tait: *Finitism.* Journal of Philosophy 78 (1981) 524-546.

[66] G. Takeuti: *Proof Theory.* Second edition (North-Holland, Amsterdam, 1987).

[67] M. Toppel: *Complexities of Proof-Theoretical Reductions.* (PhD thesis, University of Leeds, 2016). Online accessible at: http://etheses.whiterose.ac.uk/16456/