A New Model Construction by Making a Detour via Intuitionistic Theories II: Interpretability Lower Bound of Feferman's Explicit Mathematics T_0

Kentaro Sato[a,1]

[a]Institut für Informatik und angewandte Mathematik, Universität Bern, Neubrückstrasse 10, 3012 Bern, Switzerland

Abstract

We partially solve a long-standing problem in the proof theory of explicit mathematics or the proof theory in general. Namely, we give a lower bound of Feferman's system T_0 of explicit mathematics (but only when formulated on classical logic) with a concrete interpretation of the subsystem $\Sigma^1_2\text{-}\mathsf{AC} + (\mathsf{BI})$ of second order arithmetic inside T_0. Whereas a lower bound proof in the sense of proof-theoretic reducibility or of ordinal analysis was already given in 80s, the lower bound in the sense of interpretability we give here is new.

We apply the new interpretation method developed by the author and Zumbrunnen (2015), which can be seen as the third kind of model construction method for classical theories, after Cohen's forcing and Krivine's classical realizability. It gives us an interpretation between classical theories, by composing interpretations between intuitionistic theories.

Keywords: Feferman's explicit mathematics, making a detour via intuitionistic theories, interpretability, tree representation of sets, extensional realizability
2010 MSC: (Primary) 03B30, 03F25, 03F50, (Secondary) 03B40, 03D60, 03E30

1. Introduction

The present author and Zumbrunnen [37] have developed a new method to construct an interpretation between classical theories (i.e., theories based on classical logic) by composing interpretations between intuitionistic theories (theories based on intuitionistic logic, which of course include classical theories). In the (modified) sense of Visser's *miniature model theory*, this can be seen as the third kind of model construction method for classical theories, after Cohen's forcing method and Krivine's classical realizability method. Whereas the present author and Zumbrunnen [37] have applied this method in order to give lower bounds to a new family of theories introduced there, here in the present paper we apply the method to solve, partially, a long-standing problem in the proof theory of explicit mathematics or the proof theory in general.

Proof theory originally aimed to obtain, on a finitistic ground, consistency proofs of mathematical frameworks, i.e., axiomatic systems in which (a significant part of) mathematical practice can be formalized. Since Gödel's second incompleteness theorem had shown that this is impossible (with the suitable meaning of "finitistic ground"), the aim has become reducing the consistency of a mathematical framework to another. This type of proof theory is sometimes called *reductive proof theory*. For this purpose, *ordinal analysis* has been playing a major role, since Gentzen reduced the consistency of Peano arithmetic to the transfinite induction up to ε_0 (the scheme consisting of the transfinite induction for any primitive recursive predicate along any ordinal below ε_0) and therefore to any theory which proves the transfinite induction up to ε_0. Rathjen [33, Subsection 3.2], a leading figure of current proof theory, listed those results in reductive proof theory that are obtained by ordinal analysis, among which is the proof-theoretic equivalence between $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$, KPi and Feferman's system T_0 of explicit mathematics. Unlike the other results listed there, he explicitly mentions that "no proof of the above result has been found that doesn't use ordinal representation".

Email address: `sato@iam.unibe.ch` (Kentaro Sato)
[1]The research is partially supported by Swiss National Science Foundation.

Showing the proof-theoretic equivalence between these three theories without ordinal analysis has been a long-standing problem (as Rathjen's remark above; see also [12, p.12]), especially in explicit mathematics community. Since Feferman [8, Part V] gave a relative interpretation of T_0 in $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ in 70s soon after he introduced T_0 and Jäger [17, Section 8] in 80s showed the equivalence between $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ and KPi by giving concrete mutual relative interpretations, what remains is to give, without ordinal analysis, a reduction of $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ to T_0. In this paper, we solve this long-standing problem partially. By 'partially', we mean the reduction to T_0 formulated on classical logic, while Feferman originally formulated it on intuitionistic logic and Rathjen [33] apparently meant the intuitionistic version by T_0. Nonetheless, nowadays systems of explicit mathematics are formulated more often on classical logic and, moreover, recently Feferman [11] himself states that explicit mathematics formulated on classical logic can play the role which he originally intended explicit mathematics to play. Thus we could claim that the problem remains significant only for the classical version of T_0, at least in the context of present-day explicit mathematics.

1.1. Explicit mathematics

Explicit mathematics was introduced in 70s by Feferman [6, 7, 8] to provide a uniform framework for many kinds of mathematics, e.g., *Bishop-style constructive analysis*, *Russian recursive constructive mathematics*, *Weyl-style predicative mathematics*, *Borelian mathematics*. Indeed, Feferman gave the ways how mathematical arguments formalized in explicit mathematics can be interpreted within any of these kinds of mathematics. Since Bishop-style constructive analysis has been considered to be weakest among them, explicit mathematics has sometimes been considered also as a formal framework for Bishop-style constructive analysis.

Although he [8, Part I. Section 6] mentioned another important kind of mathematics, *Brouwer-style intuitionistic mathematics*, he seemed to exclude this from his scope, saying

> Brouwer's analysis based on f.c.s. [free choice sequence] has been studied in various logical formalisms by Kleene, Vesley, Kreisel, Troelstra, van Dalen and others (cf. Troelstra 1977a 1977b [[42] and [43]] for references). Various parts of this have taken settled and coherent form (and have, incidentally, been shown consistent). But efforts to treat the most general concept of f.c.s. have not yet had a convincing outcome. For mathematicians, Brouwer's theory has remained a curiosity; it has largely been of interest to logicians. Moreover, the concepts are rather special to analysis and topology and seem to have little to do with other parts of mathematics. Historically, the actual development of intuitionistic mathematics got hung up around analysis because of the need to clarify Brouwer's ideas there. ([8, pp.168-169])

Indeed, he did not give a way how arguments in explicit mathematics can be interpreted in this kind of mathematics.

Since its born in 70s, explicit mathematics has been among such families of mathematical frameworks that are of interest in proof-theoretic researches, along with *subsystems of second order arithmetic*, subsystems and supersystems of *Kripke-Platek set theory* KP, of *Martin-Löf type theories*, and of *constructive Zermelo-Fraenkel set theory* CZF.

The main features of explicit mathematics are as follows. First, it has the applicative nature, which allows us to treat operations directly (unlike in set theory where we have to encode $A[f(x)]$ by $\exists y($"$y = f(x)$"$\wedge A[y])$ with quantifiers). This is given by the structure of *non-total combinatory algebra* on individuals, which is incorporated in the formulation of explicit mathematics based on the so-called *applicative theory*, and which entails that any individual is a code of some partial operation on individuals. As this yields the machinery of *untyped λ-calculus*, we can treat all the "computable" partial operations, with a suitable sense of "computability". Second, it has the named type structure, namely any type, a formal entity intended to denote a collection of individuals, is named by an (but not necessarily unique) individual. Through this naming machinery, we can treat operations on types. For both the features, *intensionality* and *non-totality* are essential: different individuals might code or name an identical (in the extensional sense) operation or type; and there always exist some individuals coding non-total operations and those naming no type.

Feferman introduced various systems of explicit mathematics, among which are $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{T}_0\!\upharpoonright$, $\mathsf{T}_0\!\upharpoonright + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and T_0 (although he employed different notations for these systems, except T_0), and he [8, Part V] identified the proof-theoretic strengths of all these systems except T_0, by giving proof-theoretic equivalence results with subsystems of second order arithmetic. He only made a conjecture on the proof-theoretic strength of T_0, which was proven to be true later by Jäger [16] by ordinal analysis. Among these five systems, only for T_0 no proof of the proof-theoretic equivalence has previously been known that is not based on ordinal analysis[2], while it seems almost obvious from Feferman's notation that he considered T_0 as the most important among the systems of explicit mathematics.

As mentioned before, nowadays these systems are more often defined on classical logic. On this line, some extensions by *small large cardinal notions* have been introduced and investigated in Marzetta [28], Strahm [39, 40], Jäger, Kahle and Studer [19], Jäger and Studer [24], Jäger and Strahm [21, 23] and Jäger [18], as similar extensions of Kripke-Platek set theory, of constructive Zermelo-Fraenkel set theory and of Martin-Löf type theory have been considered in the right direction to strengthen these mathematical frameworks and also to develop further proof-theoretic investigations. Also for the extensions by monotone inductive definition, another direction of development, Takahashi [41], Rathjen [30, 31, 32, 35] and Glaß, Rathjen and Schlüter [14] all formulated the systems on classical logic, though Rathjen [32] mentioned extending the result for the intuitionistic version as a further problem which was treated by Tupailo [46]. Thus we could say that explicit mathematics in the present-day context is formulated almost always on classical logic.

1.2. Foundational issue on the logic on which explicit mathematics should be formulated

As mentioned above, explicit mathematics was originally designed as a uniform framework of many kinds of mathematics including constructive ones, and especially as a formal framework for Bishop-style constructive analysis. Thus many people consider that explicit mathematics must be formulated on intuitionistic logic for this original purpose, and this seems to be the reason why Feferman introduced and considered the systems on intuitionistic logic in the early ages of explicit mathematics.

However, even for this original purpose, Feferman himself states that explicit mathematics formulated on classical logic can play the intended role. Actually he explicitly says as follows.

> From this reading of Bishop style constructive analysis, I was led to introduce an axiomatic system T_0 based on classical logic in which all his work could be directly formalized. ([11, p.3])
>
> ... there is no need to restrict to intuitionistic logic in the development of Bishop's approach and in any case we want to have classical logic as our basic system of reasoning throughout in order to deal in a common way with constructive, predicative and descriptive mathematics. ([11, p.5])

As a reason for this conclusion (or as the content of "this reading of Bishop style constructive analysis"), he mentions as follows.

> Though Bishop agreed with the Brouwerians that one should restrict oneself to reasoning in intuitionistic logic, I came to the conclusion that that was not the real reason why one could give a systematic recursive interpretation to his results. Rather, its success in that respect depends essentially on two features, one general and the other more specific. The general point is that all of Bishop's basic notions are considered without assumption of extensionality, and in that sense are intensional, although in an abstract sense. (It is that which the 'Explicit', in 'Explicit Mathematics', is intended to suggest.) In particular, operations can be interpreted directly as computational programs, or indices of partial recursive functions. But they can also be considered extensionally, thus making the basic notions a part of classical mathematics. The second, more specific, feature of Bishop's methodology that, in my view, accounted for the success of his approach, was the way he modified classical notions to incorporate certain "witnessing data"

[2]For $\mathsf{T}_0{\restriction}$ and $\mathsf{T}_0{\restriction}+(\mathbb{L}\text{-}\mathsf{I_N})$, Feferman [8, Part V] used the proof-theoretic equivalence between $\mathsf{ID}_{<\nu}$ and $\mathsf{ID}^i_{<\nu}(\mathfrak{O})$ for $\nu = \omega, \varepsilon_0$, which is also in the list of results obtained by ordinal analysis from Rathjen [33]. As footnoted by Rathjen himself, however, Sieg obtained the result without use of ordinal analysis, although "his approach is still proof-theoretic as it employs cut-elimination for infinitary derivations". Moreover, if the systems of explicit mathematics are formulated on classical logic as nowadays more often done, the equivalence between $\mathsf{ID}_{<\nu}$ and $\mathsf{ID}^i_{<\nu}(\mathfrak{O})$ is no longer needed.

that is implicitly carried along in proofs. Together, his notions and results may be considered to be a refinement of classical mathematics that at the same time admits of a constructive interpretation in recursive form. Since Bishop's redevelopment of analysis is simply a part of classical mathematics, that is another way in which it diverges significantly from Brouwerian intuitionism. Brouwer treated real numbers as "choice sequences" of rational numbers, of which one would only have a finite amount of information at any time, and functions of real numbers would thus be recast in terms of functions of choice sequences. From this, Brouwer was led to the theorem that every function on a closed interval is continuous, patently contradicting classical analysis. Bishop's approach, by contrast, admits dealing with discontinuous (partial) functions on the real numbers in his theory of measure. ([11, pp.2-3])

Although the present author is not ready to take a stand in support of or against Feferman's interpretation of Bishop-style constructive analysis, what is important is that the founder of explicit mathematics, Feferman, himself claims so. Therefore there seems to be no foundational motivation to consider explicit mathematics formulated on intuitionistic logic.

Of course, there are some technical reasons to do that (even though almost all the recent technical results on explicit mathematics are for the classical versions as seen above): for example, the proof theoretic strengths of some intuitionistic systems have been established by interpreting intuitionistic T_0 (e.g., Griffor and Rathjen [15]). For this reason, there seems to be no consensus among experts with which logic the long-standing problem should be solved, or rather this paper sheds light on the lack of consensus.

1.3. Reduction vs. interpretation

As mentioned before, the proof-theoretic equivalence between $\Delta^1_2\text{-}\mathsf{CA}+(\mathsf{BI})$, KPi and T_0 has already been obtained by ordinal analysis. This result actually states more than the reducibility of the consistencies of them to one another: they prove the same arithmetical sentences (if T_0 is formulated on classical logic).

In general, a theory T_1 is said to be *proof-theoretically reducible* to another T_2 for a class Φ of formulae if there is a primitive recursive function f on finite strings (in the sense of a fixed Gödel numbering) such that

(Φ-Red) for any formula $F \in \Phi$, if x is a T_1-proof ending in F, then $f(x)$ is a T_2-proof ending in F.

Of course, here we have to assume that Φ is contained in both the languages of T_1 and of T_2. Generally this implies the Φ-conservation of T_1 over T_2 (i.e., if $T_1 \vdash F$ then $T_2 \vdash F$ for $F \in \Phi$), and as far as Φ contains $\perp$ it implies the reduction of consistency. T_1 and T_2 are said to be *proof-theoretically equivalent* if T_1 is proof-theoretically reducible to T_2 and vice versa.

If we accept the distinction between "real" and "ideal" notions in mathematics, taking Φ as the set of those formulae concerning only "real" objects (sometimes identified with arithmetical formulae) or observable in the "real" world (with Π^0_1 or Π^0_2), we can consider that Φ-conservation of T_1 over T_2 states that T_1 is not stronger than T_2, as a fictional tool to talk about the "real" objects or "real" world, or, in other words, the "ideal" objects or world described in T_2 have at least the same power as those in T_1. From this viewpoint, it is not unnatural that T_1-theorems outside Φ have no meanings in T_2, but only those in Φ.

On the other hand, the notion of interpretation does not concern the distinction between "real" and "ideal" objects. Although we do not define the general notion of interpretation, an interpretation of T_1 in T_2 at least provides us a primitive recursive transformation f of T_1-proofs to T_2-proofs satisfying:

(Preservation of proof structure) if x is a subproof of y, then $f(x)$ is a subproof of $f(y)$.

If this f satisfies additionally the condition (Φ-Red) up to the provable equivalence, the interpretation is called Φ-*preserving*. Since the second condition requires the transformation to preserve the structure of proofs, and since we do not need any restriction on the conclusions of proofs, T_2 can simulate *any* mathematical practice formulated in T_1 as a whole. In this sense, interpretability is stronger than proof-theoretic reducibility, and the *interpretability strength* could be said a strength defined from the viewpoint that how large part of mathematical practice can be formulated, whereas the proof-theoretic strength is the power of the fiction as an investigating tool on the "real" fragment.

It is true that interpretability and reducibility coincide, if all the extensions (within the same language) of the theory are reflexive (i.e., the consistency of any finite fragment is provable in the theory), as has been well known (see, e.g., Lindström [27, Theorem 8.8]), with the narrow sense of interpretation (i.e., in these results we can require the interpretations to preserve the logical connectives, as opposed to the interpretations we will consider in the present paper). However, this fact does not hold in non-reflexive cases, and even in the reflexive case the obtained interpretation satisfies only (Π^0_1-Red), but not necessarily (Φ-Red) for relevant Φ.

From the perspective based on the difference between reducibility and interpretability, we will obtain new results even for those subsystems of T_0 whose proof-theoretic strengths have previously been established without use of ordinal analysis, namely $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{T}_0\!\restriction$ and $\mathsf{T}_0\!\restriction + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$ in our notation. Especially, the reducibility of Σ^1_1-AC_0 to $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$ was already given by Feferman from the beginning of explicit mathematics, but it was via the reducibility of Σ^1_1-AC_0 to ACA_0 which does not tell how to simulate the axiom of choice (especially those inside cut-non-free proofs) within ACA_0 or $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$. With the method developed by the present author and Zumbrunnen [37], we will give interpretations of subsystems of second order arithmetic in the systems of explicit mathematics, and this automatically tells how to simulate the axiom of choice in explicit mathematics. As expected, for this simulation, the join operator $\mathbf{j}$ of explicit mathematics will play the crucial role.

1.4. Making a detour via intuitionistic theories

The method developed by the present author and Zumbrunnen [37] is to interpret classical theories in classical theories but via non-classical intuitionistic theories. Quite interestingly, in spite of this intuitionistic nature, this method does not work for intuitionistic theories so well, as explained in [37, Subsection A.4] (and this is why we have devoted one subsection for the argument on the logic on which explicit mathematics should be formulated). As mentioned there the resulting interpretation can be seen as a composition of

- Gödel-Gentzen's negative interpretation,
- intuitionistic forcing interpretation (a straightforward formalization of Kripke semantics of intuitionistic logic) and
- realizability interpretation (a straightforward formalization of Brouwer-Heyting-Kolmogorov semantics of intuitionistic logic).

As explained also there, if seen as a model construction, this is the only way of constructing models of classical theories, only besides

- the logically-trivial ones (those in which the interpretations of logical connectives are defined trivially),
- Cohen's classical forcing method (which can be seen as the composition of a negative interpretation and intuitionistic forcing interpretation) and
- Krivine's classical realizability method (which can be seen as the composition of a negative interpretation and realizability interpretation).

Though we refer to Sato and Zumbrunnen [37, Subsections 1.4 and 1.5] for the details of the significance and the novelty of this method, the author would like to repeat that *this demonstrates the utility of extending the scope to intuitionistic theories even for the investigation of classical theories.*

Sato and Zumbrunnen [37] applied this method to interpret Kripke-Platek set theories in variants of operational set theory, especially to interpret the collection scheme with the replacement operator $\mathbb{R}$. The basic idea is that first we interpret the intuitionistic collection scheme with the operator $\mathbb{R}$ via realizability interpretation at the cost of classical logic and then we interpret the classical collection scheme with the intuitionistic one via negative and forcing interpretations. The second part was a modification of the method developed by Avigad [3] in order to interpret classical theories in the corresponding intuitionistic theories.

Similarly we can interpret the intuitionistic Σ^1_1 axiom of choice scheme with the join operator $\mathbf{j}$ via realizability interpretation and then the classical Σ^1_1 axiom of choice scheme with the intuitionistic one, as Avigad [3] applied his method also to obtain the interpretability of classical Σ^1_1-AC in intuitionistic Σ^1_1-AC. However, Avigad's method (and hence the modified method by Sato and Zumbrunnen [37]) does not work well for Σ^1_2 axiom of choice. In order to reduce the complexity from Σ_2 to Σ_1 (more precisely, Σ^1_2 axiom of choice to Σ_1 collection), we first interpret a family of set theories in systems of explicit mathematics in Section 4 and then apply the method developed by Sato and Zumbrunnen [37] in Sections 5 and 6. As is well known, in the presence of foundation or regularity axiom, Σ_1 in set theory corresponds to Σ^1_2 in second order arithmetic. The family of set theories designed for this purpose is called *weak explicit set theory with non-set operators*, which will be explained in the next subsection.

1.5. Weak explicit set theory with non-set operators

We will introduce a family of set theories, called weak explicit set theory with non-set operators, for the aforementioned purpose. These are variants of *weak explicit set theory*, introduced in Sato and Zumbrunnen [37], for which the method developed there works well. Weak explicit set theory is defined from operational set theory by replacing the axioms of truth functions on Δ_0 formulae and of separation operator with the axioms of Gödel operations. Our weak explicit set theory with non-set operators is defined further by allowing some operators not to be sets with a new predicate for set-hood and the equality as operators (the same as the absolute equality $=$) to differ from the equality as sets (which will be denoted by $\approx$).

The first connection to explicit mathematics is that these theories are interpreted in systems of explicit mathematics via tree representation of sets, as we will show in Section 4. This representation naturally allows non-set objects to exist which of course code some operations and the absolute equality to differ from the equality as sets, which is interpreted as the extensional equality between trees.

The second is the inheritance of the features of explicit mathematics. As previously mentioned, the important features of explicit mathematics are intensionality and non-totality both for the operational and typing machineries. While both operational set theory, introduced by Feferman [9, 10], and weak explicit set theory, introduced in Sato and Zumbrunnen [37], inherit these two features for the operational machinery, they do not inherit them for the typing one, for they contain the axiom of extensionality for sets and they assume that all formal objects are sets (where sets substitute types). Moreover, because of the axiom of extensionality, the distinction between operations and set-theoretic functions is dissolved, as proved, e.g., in [37, Proposition 13]. Our weak explicit set theory with non-set operators, on the other hand, inherits all of these features of explicit mathematics, since it does not contain the axiom of extensionality for sets (even in the sense of the weaker equality $\approx$) nor assume that all objects are sets. Because of the absence of the axiom of extensionality, the distinction between operations and set-theoretic functions makes sense in this framework. Indeed, many ontological differences between explicit mathematics and operational set theory investigated in Jäger and Zumbrunnen [25] do not apply to our weak explicit set theory with non-set operators.

For these reasons, we could claim that *weak explicit set theory with non-set operators is a proper set-theoretic counterpart of explicit mathematics*, whereas in the literature operational set theory is sometimes considered to be so. Nonetheless, in the present paper, we use the systems of this new family just as auxiliary systems and the significance of the lower bounds of them, for our purpose, is that they give us the lower bounds of systems of explicit mathematics.

1.6. Extensional realizability

Since we modify the weak explicit set theory, the results obtained in Sato and Zumbrunnen [37] do not apply directly. The part we have to modify is the realizability interpretation of variants of intuitionistic Kripke-Platek set theory in weak explicit set theory. For we have to interpret the equality $=$ of the set theory as $\approx$, which differs from the absolute equality $=$ of weak explicit set theory with non-set operators.

Let us see the need of modification more closely. To obtain the realizability of the collection scheme, we have to construct an operation which assigns both a set b and a realizer of $(\forall x \in a)(\exists y \in b)A[x, y]$ to a realizer of the premise $(\forall x \in a)\exists y A[x, y]$. Although any realizer of $(\forall x \in a)\exists y A[x, y]$, in the sense of the usual realization relation, yields an operator c such that $A[x, cx]$ is realizable for all $x \in a$, nothing guarantees that this operator c is extensional, i.e., $cx_1 \approx cx_2$ whenever $x_1 \approx x_2$. Since the replacement operator $\mathbb{R}$ applies only to extensional operations, we have to modify the realization relation so that any realizer of $(\forall x \in a)\exists y A[x, y]$ yields an *extensional* operator c such that $A[x, cx]$ is realizable for all $x \in a$.

The basic idea of the modification we will make is to introduce, for each formula A of the realized theory, a relation $\sim_A$ which is intended to be the equality among the realizers of A, and to require all the relevant operations to be extensional with respect to these $\sim_A$'s. Particularly, the equality $\sim_{\exists x A[x]}$ for the realizers of $\exists x A[x]$ is defined as the conjunction of $\approx$ between the witnesses and $\sim_{A[x]}$ between the realizers of $A[x]$ with the witness substituted to x; and the realization relation for $\forall x B[x]$ requires realizers to be extensional with respect to $\approx$ and $\sim_{B[x]}$. With this, we can guarantee that the realizers of $(\forall x \in a)\exists y A[x, y]$ yield extensional operations. Moreover, this does not affect the expected properties of realizability interpretation, for example, the closure under intuitionistic inferences. Similar "extensional" realization relations appeared in, e.g., Troelstra [44], Griffor and Rathjen [15], and Oosten [29] (where it is called **e**-realizability).

From the viewpoint that the realizability machinery is a "miniature" of Martin-Löf type theory, the use of $\sim_A$ can be seen as a use of identity type. For the details of this point and for the comparison with the realizability in Tupailo [45], see the comments at the end of Subsection 5.2.

1.7. Well-foundedness as a predicate vs. well-foundedness as an operation

As a byproduct of our investigation, we can easily design a system $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$ of explicit mathematics whose strength is the same as ID_1, the strength of so-called *generalized predicativity.* Our method can also establish the mutual interpretability of this system with Kripke-Platek set theory, the most popular among systems of this strength. The basic idea is to replace the so-called inductive generation operator **ig** from T_0 by transfinite induction scheme along those relations designated by a predicate $\mathfrak{I}$ which is intended to represent the "real" well-foundedness. Since the operator **ig** assigns the well-founded (accessible) parts, $\mathfrak{I}$ can be characterized as the image of the operator. Thus, we could say that $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$ is defined with the notion of well-foundedness as a predicate whereas T_0 is with that as an operation. The final result on the strengths of these systems clarifies the difference between these two approaches: well-foundedness as an operation claims one level higher in the projective hierarchy than well-foundedness as a predicate.

This point might become even clearer if we apply our technique to the applicative theories, which do not have the type structure but the so-called type-2 functionals instead. As will be discussed briefly in Subsection 4.5, our technique yields also lower bounds of the following theories: $\mathsf{BON}(\boldsymbol{\mu})$, from Feferman and Jäger [13], which has the so-called non-constructive μ-operator $\boldsymbol{\mu}$; and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$, from Jäger and Strahm [22], which has both $\boldsymbol{\mu}$ and Suslin operator $\boldsymbol{S}$. Since Suslin operator tells us whether the input encodes a well-founded relation or not, this is a more direct operational analogue of the well-foundedness predicate. We could obtain the same contrast between $\mathsf{BON}(\boldsymbol{\mu})$ with transfinite induction along those relations designated by the well-foundedness predicate and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$.

This may explain why it is in general harder in the framework of explicit mathematics to design a subsystem of the strength of generalized predicativity: explicit mathematics is a framework in which new notions are added as operations rather than as predicates. Nonetheless, it is not impossible, as another system of the same strength was introduced by Kahle and Studer [26].

1.8. Outline of the paper

In the next section, Section 2, we recall the formal definitions of some systems of explicit mathematics. Our formulation is somehow different from the standard ones, but the equivalence of these formulations can be seen easily. Since the readers are assumed to be familiar with subsystems of second order arithmetic and with variants of Kripke-Platek set theory, we do not give the detailed definitions of them (except Definitions 50 and 51). The readers not familiar with these topics can refer to Simpson [38] for the former and both Barwise [4, Chapter I] and Sato and Zumbrunnen [37, Section 2] for the latter. Nonetheless, we recall that $\Sigma^1_2\text{-}\mathsf{CA}_0$ has been known to be identical (i.e., not only proof-theoretically but also logically equivalent) to $\Delta^1_2\text{-}\mathsf{CA}_0$ as shown, e.g., in Simpson [38, Theorem VII.6.9], and accordingly that both $\Sigma^1_2\text{-}\mathsf{CA}$ and $\Sigma^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ are identical with $\Delta^1_2\text{-}\mathsf{CA}$ and $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ respectively.

In Section 3, we give formal definitions of systems of weak explicit set theory with non-set operators, and in the following section, Section 4, we interpret them in systems of explicit mathematics with interpretations $\natural$ and $\sharp$, both of which are based on the tree-representation of sets. While both the language $\mathbb{L}$ of explicit mathematics and that $\mathcal{L}_\in$ of set theory can be seen as extensions of the language $\mathcal{L}_2$ of second order

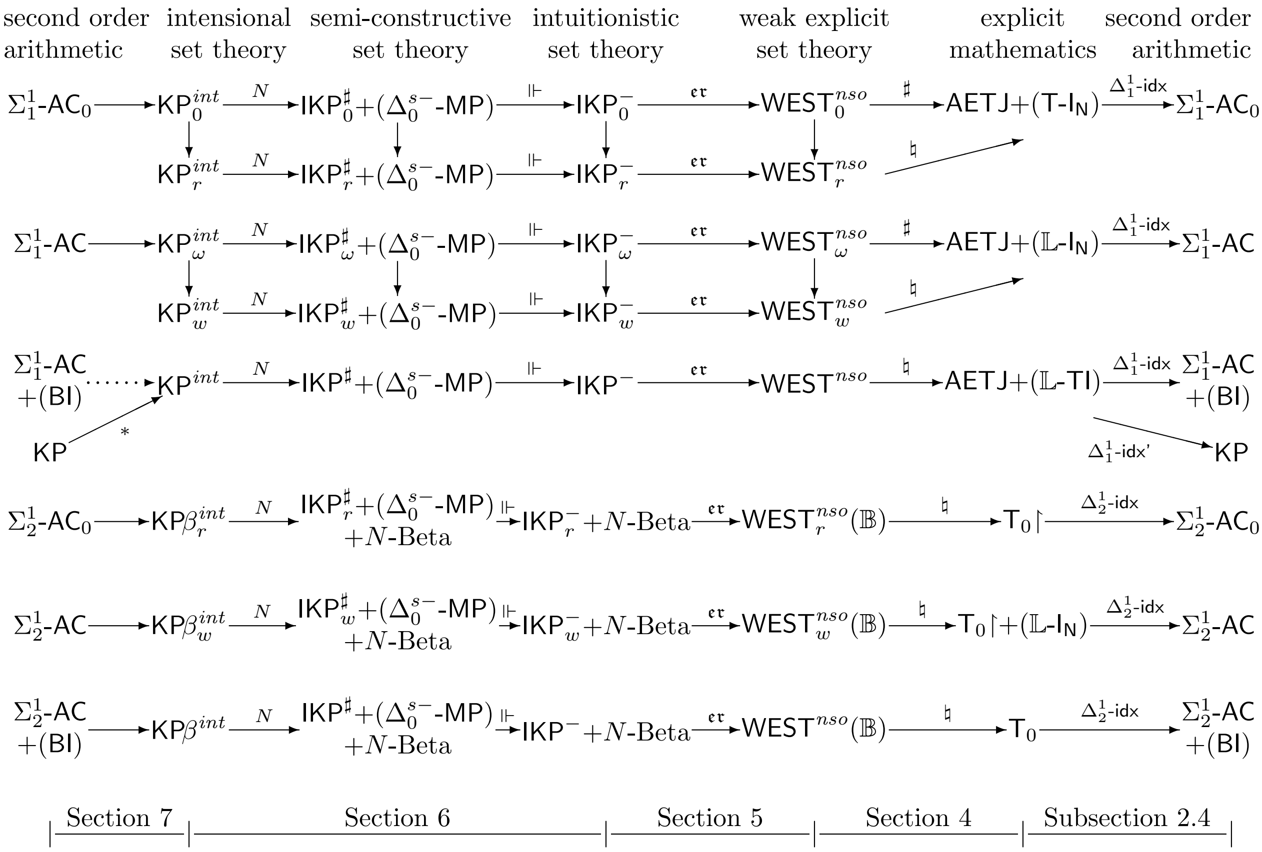

Figure 1: The relation between interpretations and theories

arithmetic, we will see that these interpretations preserve $\mathcal{L}_2$-formulae. It will also be mentioned in Subsection 4.5 that, for these interpretations, the non-constructive μ-operator $\boldsymbol{\mu}$ can play the role of the type structure of explicit mathematics.

In Section 5, we introduce the extensional realizability interpretation $\mathfrak{er}$, which interprets, in weak explicit set theory with non-set operators, intuitionistic Kripke-Platek set theory with the Δ_0 separation scheme restricted to negative formulae. We will see that this interpretation preserves negative formulae.

In Section 6, we summarize Avigad's method but modified in Sato and Zumbrunnen [37, Sections 5 and 6] for intuitionistic Kripke-Platek set theory with the Δ_0 separation scheme restricted to negative formulae, and apply it to interpret intensional variants of Kripke-Platek set theory, i.e., Kripke-Platek set theory without the axiom of extensionality. This is a combination of a forcing interpretation $\Vdash$ and Gödel-Gentzen's negative interpretation N. We can check that, according to this interpretation, any Π_1-formula is interpreted as a classically equivalent negative Π_1-formula.

In Section 7, we check that some results known for Kripke-Platek set theory survive in the absence of extensionality, and obtain the results that the relevant subsystems of second order arithmetic are contained in the corresponding variants of intensional Kripke-Platek set theory.

In Section 8, we summarize these interpretability results to obtain the final results.

Figure 1 shows which interpretation interprets which theory in which theory, where the dotted arrow means a proof-theoretic reducibility. Referring to this figure in each section the readers could have some idea on which part of the whole process is executed.

2. Explicit Mathematics

2.1. The language $\mathbb{L}$ of explicit mathematics

Definition 1 (Language $\mathbb{L}$)**.** The language $\mathbb{L}$ of explicit mathematics is a two-sorted language, whose sorts are called individual and type. Lower case Latin letters $a, b, c, f, g, h, u, v, w, x, y, z$ etc. possibly with subscripts denote individual variables and capital ones X, Y, Z, U, V, W etc. possibly with subscripts denote type variables.

$\mathbb{L}$ has a binary function symbol $\circ$ for individual (application of individual to individual), unary relation symbols $\downarrow$ (definedness), N (being a natural number) all for the individual sort and $\Im$ (well-foundedness) for the type sort; and binary relation symbols $=$ (equality) for the individual sort, $\in$ (membership) between the individual and type sorts and $\Re$ (naming representation) between the individual and type sorts.

Furthermore $\mathbb{L}$ has the following individual constants: $\mathbf{k}$ and $\mathbf{s}$ (K and S combinators); $\mathbf{p}$, $\mathbf{p}_0$ and $\mathbf{p}_1$ (pairing and projections); 0 (zero), $\mathbf{s}_\mathsf{N}$ (successor), $\mathbf{p}_\mathsf{N}$ (predecessor), $\mathbf{d}_\mathsf{N}$ (definition by numerical cases) and $\mathbf{r}_\mathsf{N}$ (recursor); and **nat** (name of the type of natural numbers), **id** (name of the type of diagonal pairs), **co** (name of the complements), **un** (name of binary unions), **dom** (name of domains), **inv** (name of inverse images), **j** (name of joins), and **ig** (name of inductive generations).

For readability, $\circ(s,t)$ is often denoted by $s \circ t$ or simply st, and both $st_1...t_n$ and $s(t_1,...,t_n)$ denote $(...(st_1)...)t_n$. Furthermore, $\langle s,t\rangle$ denotes $\mathbf{p}(s,t)$ and 1 denotes $\mathbf{s}_\mathsf{N}(0)$. Also we introduce the following abbreviations:

$$
\begin{aligned}
s \simeq t &:\equiv s\downarrow \vee\, t\downarrow \to s = t & s \in \mathsf{N} &:\equiv \mathsf{N}(s)\\
X = Y &:\equiv \forall x(x \in X \leftrightarrow x \in Y) & X \subseteq Y &:\equiv \forall x(x \in X \to x \in Y)\\
s \,\dot{\in}\, t &:\equiv \exists X(\Re(t,X) \wedge s \in X) & s \,\dot{\subseteq}\, t &:\equiv \exists X, Y(\Re(s,X) \wedge \Re(t,Y) \wedge X \subseteq Y)\\
\Re(s) &:\equiv \exists X \Re(s,X) & \Im(s) &:\equiv \exists X(\Re(s,X) \wedge \Im(X))\\
\Re(\vec{s},\vec{X}) &:\equiv \Re(s_1,X_1) \wedge ... \wedge \Re(s_n,X_n) & &
\end{aligned}
$$

where $\vec{s}$ and $\vec{X}$ are the sequences $s_1,...,s_n$ and $X_1,...,X_n$ respectively. Also, we use the usual set-theoretic notations for types, e.g., $A[\{x \in U : B[x]\}]$ abbreviates $\exists X(A[X] \wedge \forall x(x \in X \leftrightarrow x \in U \wedge B[x]))$, $X \cap Y$ denotes $\{x : x \in X \wedge x \in Y\}$, and $X \times Y$ denotes $\{\langle x,y\rangle : x \in X \wedge y \in Y\}$. The abbreviation $A[\{x \in U : B[x]\}]$ might cause a confusion if there is no type X such that $\forall x(x \in X \leftrightarrow x \in U \wedge B[x])$. However, if $B[x]$ is elementary, Lemma 6 tells us that there is no such danger.

Definition 2 (elementary $\mathbb{L}$-formula)**.** An $\mathbb{L}$-formula is called *elementary* if it contains no occurrences of $\Im$, $\Re$ nor type quantifiers.

Note that elementary formulae might contain the definedness predicate $\downarrow$ and type variables as parameters.

Remark 3. The language $\mathcal{L}_2$ of second order arithmetic can be embedded into $\mathbb{L}$ in an obvious manner, namely the first order quantifiers are interpreted as quantifiers over N and the second order quantifiers are interpreted as type quantifiers. Fixing this embedding, we will consider $\mathcal{L}_2$ (and hence the language $\mathcal{L}_1$ of first order arithmetic) as a sublanguage of $\mathbb{L}$.

2.2. Applicative theory AET of elementary typing

The most basic system of explicit mathematics is the applicative theory AET of elementary typing. We basically follow the formulation given by Jäger and Zumbrunnen [25], but with some revisions convenient for our purpose.

Definition 4 (Theory AET)**.** The theory AET (applicative theory of elementary typing) is an $\mathbb{L}$-theory whose underlying logic is a classical logic of partial terms (due to Beeson [5]; which treats $\downarrow$ as a logical symbol) with equality for individuals, and whose non-logical axioms are the following.

The so-called *applicative axioms* are standard axioms about the combinators and operations on natural numbers:

(I.1) $\mathbf{k}ab = a \wedge \mathbf{s}ab\downarrow \wedge \mathbf{s}abc \simeq (ac)(bc)$,

(I.2) $\langle a,b\rangle\downarrow \wedge \mathbf{p}_0\langle a,b\rangle = a \wedge \mathbf{p}_1\langle a,b\rangle = b$,

(I.3) $0 \in \mathsf{N} \wedge (a \in \mathsf{N} \to \mathbf{s}_\mathsf{N}(a) \in \mathsf{N} \wedge \mathbf{s}_\mathsf{N}(a) \neq 0 \wedge \mathbf{p}_\mathsf{N}(\mathbf{s}_\mathsf{N}(a)) = a)$,

(I.4) $\mathbf{p}_\mathsf{N}(0) = 0 \wedge (a \in \mathsf{N} \wedge a \neq 0 \to \mathbf{p}_\mathsf{N}(a) \in \mathsf{N} \wedge \mathbf{s}_\mathsf{N}(\mathbf{p}_\mathsf{N}(a)) = a)$,

(I.5) $a \in \mathsf{N} \wedge b \in \mathsf{N} \to (a = b \to \mathbf{d}_\mathsf{N}(u,v,a,b) = u) \wedge (a \neq b \to \mathbf{d}_\mathsf{N}(u,v,a,b) = v)$,

(I.6) $(\forall x_0,\dots,x_n \in \mathsf{N})(g(x_0,\dots,x_n) \in \mathsf{N}) \wedge (\forall x_0,\dots,x_n,x_{n+1},x_{n+2} \in \mathsf{N})(f(x_0,\dots,x_n,x_{n+1},x_{n+2}) \in \mathsf{N})$
$\to (\forall x_0,\dots,x_n,x_{n+1} \in \mathsf{N})\begin{pmatrix} \mathbf{r}_\mathsf{N}(f,g)(x_0,\dots,x_n,x_{n+1}) \in \mathsf{N} \wedge \mathbf{r}_\mathsf{N}(f,g)(x_0,\dots,x_n,0) = g(x_0,\dots,x_n) \\ \wedge \mathbf{r}_\mathsf{N}(f,g)(x_0,\dots,x_n,\mathbf{s}_\mathsf{N}(x_{n+1})) = f(x_0,\dots,x_n,x_{n+1},\mathbf{r}_\mathsf{N}(f,g)(x_0,\dots,x_n,x_{n+1}))\end{pmatrix}$.

The axioms of the second group, the so-called *explicit representation axioms*, assert the basic rule of naming machinery:

(II.1) Every type has a name $\exists x\Re(x,X)$,

(II.2) Naming is unique $\Re(a,U) \wedge \Re(a,V) \to U = V$.

The axioms of the third group are called *uniform naming axioms*:

(III.1) natural number $\Re(\mathbf{nat}) \wedge (a \mathrel{\dot\in} \mathbf{nat} \leftrightarrow a \in \mathsf{N})$,

(III.2) identity $\Re(\mathbf{id}) \wedge (a \mathrel{\dot\in} \mathbf{id} \leftrightarrow \exists b(a = \langle b,b\rangle))$,

(III.3) complement $\Re(u) \to \Re(\mathbf{co}(u)) \wedge (a \mathrel{\dot\in} \mathbf{co}(u) \leftrightarrow \neg(a \mathrel{\dot\in} u))$,

(III.4) union $\Re(u) \wedge \Re(v) \to \Re(\mathbf{un}(u,v)) \wedge (a \mathrel{\dot\in} \mathbf{un}(u,v) \leftrightarrow a \mathrel{\dot\in} u \vee b \mathrel{\dot\in} v)$,

(III.5) domain $\Re(u) \to \Re(\mathbf{dom}(u)) \wedge (a \mathrel{\dot\in} \mathbf{dom}(u) \leftrightarrow \exists y(\langle a,y\rangle \mathrel{\dot\in} u))$,

(III.6) inverse image $\Re(u) \to \Re(\mathbf{inv}(u,f)) \wedge (a \mathrel{\dot\in} \mathbf{inv}(u,f) \leftrightarrow fa \mathrel{\dot\in} u)$.

The fourth group of axioms governs the additional relation symbol $\Im$:

(IV.1) type transfinite induction $\Im(W) \wedge \forall x(\forall y(\langle y,x\rangle \in W \to y \in X) \to x \in X) \to \forall x(x \in X)$,

(IV.2) operational transition $\Im(W) \wedge (\exists f)(\forall x,y)(\langle x,y\rangle \in R \to \langle fx,fy\rangle \in W) \to \Im(R)$;

(IV.3) progression $(\forall x(\langle x,a\rangle \in W \to \Im(W\restriction x))) \to \Im(W\restriction a)$, where $W\restriction x$ is W restricted to elements hereditarily below x, i.e., $\{\langle y,z\rangle \in W : \exists a(\exists n \in \mathsf{N})(a0 = x \wedge (\forall k < n)(\langle a(\mathbf{s}_\mathsf{N}(k)), ak\rangle \in W) \wedge an = z)\}$.

Note that some of the constants are not mentioned in the axioms of AET. These are needed in the definitions of extensions of AET, and play no role in AET.

Axiom (IV.2) guarantees that if $W = R$, i.e., $(\forall w)(w \in W \leftrightarrow w \in R)$, then $\Im(W) \leftrightarrow \Im(R)$.

In any previous formulation of explicit mathematics, the fourth group did not occur. This is added for our convenience and these axioms do not affect the strength of AET, since we can interpret our version of AET in AET without these axioms by interpreting $\Im(W)$ as

$$(\forall X)(\,(\forall x)((\forall y)(\langle y,x\rangle \in W \to y \in X) \to x \in X) \to (\forall x)(x \in X)\,).$$

The following are basic well known facts about AET.

Lemma 5. (i) For any $\mathbb{L}$-term $t[\vec{x}]$, there is an $\mathbb{L}$-term, denoted by $\lambda\vec{x}.t[\vec{x}]$, such that AET proves

$$(\lambda\vec{x}.t[\vec{x}])\downarrow \wedge (\lambda\vec{x}.t[\vec{x}])\vec{y} \simeq t[\vec{y}].$$

(ii) There is an $\mathbb{L}$-term $\mathbf{fix}$ such that AET proves

$$\mathbf{fix}(f)\downarrow \wedge \mathbf{fix}(f)(x) \simeq f(\mathbf{fix}(f),x).$$

This is a standard fact for combinatory algebra, and can be proved only by the first axiom.

Also, because of the presence of the recursor $\mathbf{r}_\mathsf{N}$, any primitive recursive function can be represented by a closed $\mathbb{L}$-term as an operation. In particular, we write $x \leq y$ for $\dot{-}(x,y) = 0$ where $\dot{-}$ is the term for the modified subtraction.

Lemma 6 (uniform elementary comprehension)**.** For any elementary formula $B[x, \vec{y}, \vec{U}]$ with at most the indicated free variables there is a closed $\mathbb{L}$-term t_B such that AET proves

$$\Re(\vec{u}, \vec{U}) \;\to\; \Re(t_B(\vec{y}, \vec{u})) \wedge \forall x(x \,\dot{\in}\, t_B(\vec{y}, \vec{u}) \leftrightarrow B[x, \vec{y}, \vec{U}]).$$

This gives us an operation which returns the name of the type $\{x : B[x, \vec{y}, \vec{U}])\}$ uniformly in the parameters. However, since the operator cannot apply to the type parameters $\vec{U}$, it applies to their names $\vec{u}$ instead as well as the individual parameters $\vec{y}$.

2.3. Join, induction, inductive generation

We consider several extensions of AET.

Definition 7 (Join)**.** AETJ is the extension of AET augmented by the axiom for join:

(III.7) join $\Re(u) \wedge (\forall x \dot{\in} u)\Re(fx) \;\to\; \Re(\mathbf{j}(u,f)) \wedge (a \,\dot{\in}\, \mathbf{j}(u,f) \;\leftrightarrow\; (\exists y, z)(a = \langle y, z\rangle \wedge y \,\dot{\in}\, u \wedge z \,\dot{\in}\, fy))$.

Definition 8 (induction schemata in explicit mathematics)**.** Type induction (T-I_N) on natural numbers and full induction ($\mathbb{L}$-I_N) on natural numbers are the following axiom schemata:

(T-I_N) $\Im(\{w : (\exists x, y \in \mathsf{N})(w = \langle x, y\rangle \;\wedge\; x < y)\})$,

($\mathbb{L}$-I_N) $A[0] \;\wedge\; (\forall x \in \mathsf{N})(A[x] \to A[\mathbf{s}_\mathsf{N}(x)]) \;\to\; (\forall x \in \mathsf{N})A[x]$ for arbitrary $\mathbb{L}$-formula $A[x]$.

Full transfinite induction ($\mathbb{L}$-TI) is the following axiom scheme:

($\mathbb{L}$-TI) $\Im(W) \;\wedge\; (\forall x)(((\forall y)(\langle y, x\rangle \in W \to A[y])) \to A[x]) \;\to\; (\forall x)A[x]$ for arbitrary $\mathbb{L}$-formula $A[x]$.

The formulation of transfinite induction might seem strange, because the field of the well-founded relation is not mentioned. However, if a relation W is well-founded on the field Y, then $W \cap (Y \times Y)$ is well-founded on the field $\{x : x = x\}$ and the transfinite induction for X along W on $x \in Y$ is equivalent to that for $X \cup \{x : x \notin Y\}$ along $W \cap (Y \times Y)$.

Definition 9 (Feferman's theories $\mathsf{T}_0\!\restriction$ and T_0)**.** $\mathsf{T}_0\!\restriction$ is the extension of AETJ augmented by (T-I_N) and

(III.8) inductive generation $\Im(R) \leftrightarrow (\exists a, r)(\Re(a) \;\wedge\; \Re(r) \;\wedge\; R = \{\langle x, y\rangle \in r : y \,\dot{\in}\, \mathbf{ig}(a, r)\})$.

T_0 is the extension of $\mathsf{T}_0\!\restriction$ augmented by ($\mathbb{L}$-TI).

T_0 was introduced in 1970s by Feferman [6, 7, 8] as the most important theory among the various systems of explicit mathematics. At the beginning it was defined on intuitionistic logic and later it is more often considered on classical logic, as mentioned in Introduction.

2.4. Feferman's Δ-index interpretation

In this subsection we recall upper bound proofs for $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$, $\mathsf{T}_0\!\restriction$, $\mathsf{T}_0\!\restriction + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and T_0, by so-called Δ-index interpretation, introduced by Feferman [8, V]. More precisely, we can interpret the first three $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$, $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$ in Σ^1_1-AC_0, Σ^1_1-AC and Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$ respectively by interpreting the types as Δ^1_1 indices, and the last three $\mathsf{T}_0\!\restriction$, $\mathsf{T}_0\!\restriction + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and T_0 in Σ^1_2-AC_0, Σ^1_2-AC and Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$ respectively by interpreting the types as Δ^1_2 indices. In both the interpretations, $\Im$ is interpreted as well-foundedness.

Definition 10 (Δ^1_n-index interpretation)**.** First we fix a natural number $\hat{c}$ for each individual constant c. For an $\mathbb{L}$-term t and a variable x not free in t, a Σ^0_1 formula $[\![t]\!](x)$ is defined as follows, where $\{\text{-}\}$ is the so-called Kleene bracket, namely $\{e\}$ denotes the partial recursive function with recursive index e:

1. for an individual constant c, $[\![c]\!](x) :\equiv x = \hat{c}$,
2. for an individual variable y, $[\![y]\!](x) :\equiv x = y$,
3. $[\![t_1 t_2]\!](x) :\equiv \exists y, z([\![t_1]\!](y) \wedge [\![t_2]\!](z) \wedge \{y\}(z) = x)$.

For $n \geq 1$, let S^1_n and P^1_n be universal Σ^1_n and Π^1_n formulae respectively (thus, for any Σ^1_n formula $B[x]$ without free variables other than x, there is a natural number e such that $\forall x(B[x] \leftrightarrow S^1_n[e,x])$). For an $\mathbb{L}$-formula A, an $\mathcal{L}_2$ formula $A^{\Delta^1_n\text{-}\mathsf{idx}}$ the Δ^1_n-index interpretation of A is defined as in Figure 2.

$$(t\downarrow)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists x([\![t]\!](x)); \qquad (x \in \mathsf{N})^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \top;$$
$$(t_1 = t_2)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists x([\![t_1]\!](x) \wedge [\![t_2]\!](x)); \qquad (t \in Y)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists x([\![t]\!](x) \wedge x \in Y);$$
$$(\Re(t,Y))^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists x([\![t]\!](x) \wedge Y = \{u : S^1_n[(x)_0, u]\} = \{u : P^1_n[(x)_1, u]\});$$
$$(\Im(W))^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \forall X(\forall x(\forall y(\langle y,x\rangle \in W \to y \in X) \to x \in X) \to \forall x(x \in X));$$
$$\bot^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \bot; \qquad (A \to B)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv (A^{\Delta^1_n\text{-}\mathsf{idx}}) \to (B^{\Delta^1_n\text{-}\mathsf{idx}}); \qquad (\forall x A[x])^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \forall x(A[x]^{\Delta^1_n\text{-}\mathsf{idx}});$$
$$(A \wedge B)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv A^{\Delta^1_n\text{-}\mathsf{idx}} \wedge B^{\Delta^1_n\text{-}\mathsf{idx}}; \quad (A \vee B)^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv A^{\Delta^1_n\text{-}\mathsf{idx}} \vee B^{\Delta^1_n\text{-}\mathsf{idx}}; \quad (\exists x A[x])^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists x(A[x]^{\Delta^1_n\text{-}\mathsf{idx}});$$
$$(\exists X A[X])^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \exists y, z, X(X = \{u : S^1_n[y,u]\} = \{u : P^1_n[z,u]\} \wedge A[X]^{\Delta^1_n\text{-}\mathsf{idx}});$$
$$(\forall X A[X])^{\Delta^1_n\text{-}\mathsf{idx}} :\equiv \forall y, z, X(X = \{u : S^1_n[y,u]\} = \{u : P^1_n[z,u]\} \to A[X]^{\Delta^1_n\text{-}\mathsf{idx}}).$$

Figure 2: Definition of Δ-index interpretation Δ^1_n-idx

The following is a well-known result in explicit mathematics, proved by Feferman [8, V], where he used so-called A-K-S Substitution Theorem to show the results concerning Δ^1_2-idx.

Theorem 11. We can chose $\hat{c}$ for each individual constant c, so that Δ^1_1-idx interprets

(i) $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I_N})$ in Σ^1_1-AC_0;
(ii) $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{I_N})$ in Σ^1_1-AC; and
(iii) $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$ in Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$,

and that Δ^1_2-idx interprets

(iv) $\mathsf{T}_0\!\upharpoonright$ in Σ^1_2-AC_0;
(v) $\mathsf{T}_0\!\upharpoonright + (\mathbb{L}\text{-}\mathsf{I_N})$ in Σ^1_2-AC; and
(vi) T_0 in Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$,

both in an $\mathcal{L}_1$-preserving way, where as proclaimed in Remark 3, we consider $\mathcal{L}_1$ as a sublanguage of $\mathbb{L}$.

Feferman [8, V] posed a question: if these (i)-(vi) are optimal. Many of them were shown by Feferman himself to be optimal, but (vi) was later shown by Jäger [16] to be optimal in the proof-theoretic sense. These results do not guarantee that there are converse interpretations, which we are giving.

In some cases, variants of Kripke-Platek set theory are better qualified for reference systems than subsystems of second order arithmetic. Actually, Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$ is less popular than KP, as a reference system of the strength of generalized predicativity. Since the language $\mathcal{L}_2$ of second order arithmetic is considered to be a sublanguage of the language $\mathcal{L}_\in$ of set theory, we can define the interpretations Δ^1_n-idx' of explicit mathematics in variants of Kripke-Platek set theory, by the same definition as Δ^1_n-idx but

$$(\Im(W))^{\Delta^1_n\text{-}\mathsf{idx}'} :\equiv (\exists f)(\mathrm{Fun}[f] \wedge \mathrm{Dom}[f,W] \wedge (\forall x,y)(\langle x,y\rangle \in W \to f'x \in f'y)),$$

where $\mathrm{Fun}[f] \wedge \mathrm{Dom}[f,W]$ means that f is a function defined on W and $f'x$ is the value of f at x (see the abbreviations defined in the next section; where $\approx$ should be replaced with $=$).

Theorem 12. Δ^1_n-idx' interprets $\mathsf{AETJ} + (\mathbb{L}\text{-}\mathsf{TI})$ in KP in an $\mathcal{L}_1$-preserving way.

3. Weak Explicit Set Theory with Non-Set Operators

Weak explicit set theory was introduced by Sato and Zumbrunnen [37], by replacing truth functions and the separation operator with Gödel operations from operational set theory without choice operator. Since the type generators in explicit mathematics play more or less the same role as Gödel operations on sets, weak explicit set theory seems easier to be interpreted in explicit mathematics than operational set theory.

However, weak explicit set theory is not handy enough. In weak explicit set theory, as well as in operational set theory, all the objects are sets, and so all the operators are sets. This seems to be a non-problematic assumption when we are working in the set theories. However, when we try to interpret (or to construct models of) these set theories, this assumption makes the construction quite tedious and complicated. Moreover, even when we are working in the set theories, this assumption plays only a few roles. For this motivation, we here introduce a version of weak explicit set theory, in which there might be non-set objects and hence non-set operators. To make it possible, we need to add a predicate for set-hood to the language of weak explicit set theory.

The relation between sets and non-sets in our situation is much simpler than in the set theory with urelements: in our situation, a non-set cannot be an element of any set, and so elements of a set are all sets.

We need two kinds of equality: one is the absolute equality which makes sense for all the objects including non-sets while the other is a weaker equality which makes sense only for sets. Also, we remove the axiom of extensionality for both the equalities, because it also make the interpretation of these theories more complex, and because it is not necessary for the interpretation of second order arithmetic in these theories.

As discussed in Subsection 1.5, these modifications make the theory inherit more features from explicit mathematics.

3.1. Languages of weak explicit set theory with non-set operators

Definition 13. The language $\mathcal{L}_\in$ of set theory is a one-sorted language with one constant symbol ω and binary relation symbols $=$ and $\in$.

Because we will also work with intuitionistic theories, we consider $\bot$ as a logical symbol, and negation is defined as $\neg A := A \to \bot$.

The bounded quantifiers $(\forall x \in a)$ and $(\exists x \in a)$ are abbreviations and are defined as usual.

Definition 14. The language $\mathcal{L}_\in^{nso}$ is the result of extending $\mathcal{L}_\in$ (which includes ω and $=$) by the constants $\mathbf{k}$, $\mathbf{s}$, $\mathbf{p}$, $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbb{R}$, $\mathbb{K}$, $\mathbb{T}$, $\mathbb{D}$, $\mathbb{U}$, $\mathbb{B}$, $\mathbb{G}_1$, $\mathbb{G}_2$, $\mathbb{G}_3$, $\mathbb{G}_4$, $\mathbb{G}_5$, $\mathbb{G}_\approx$ and $\mathbb{G}_\in$, two unary relation symbols $\downarrow$ (called the *definedness* predicate) and $\mathcal{S}$ (called *set-hood* predicate), a binary relation symbol $\approx$ (called *weak intensional equality*), and a binary function symbol $\circ$ (called *application*).

We use the abbreviation concerning $\circ$ as before: $\circ(s,t)$ is often denoted by $s \circ t$ or simply st, and both $st_1...t_n$ and $s(t_1,...,t_n)$ denote $(...(st_1)...)t_n$. However, $\langle s,t\rangle$ does not denote $\mathbf{p}(s,t)$, but the Kuratowski pair $\mathbb{D}(\mathbb{D}(s,s),\mathbb{D}(s,t))$, which is intended to denote $\{\{s\},\{s,t\}\}$. $s \simeq t$ stands for $s\downarrow \vee t\downarrow \to s = t$.

$(\forall x \in \mathcal{S})A[x]$ and $(\exists x \in \mathcal{S})B[x]$ stand for $(\forall x)(\mathcal{S}(x) \to A[x])$ and $(\exists x)(\mathcal{S}(x) \wedge B[x])$ respectively.

Definition 15. For an $\mathcal{L}_\in$ formula A, $A^\approx$ is the result of replacing all the occurrences of $=$ by $\approx$ and of quantifiers $\forall y$ and $\exists y$ by $\forall y \in \mathcal{S}$ and $\exists y \in \mathcal{S}$ respectively. A formula of $\mathcal{L}_\in^{nso}$ is called Δ_0 *formula* if it is $A^\approx[\vec{x}]$ for some Δ_0 formula $A[\vec{x}]$ in $\mathcal{L}_\in$. The classes of Δ, Π, Σ, Π_n and Σ_n formulae (for each natural number n) are defined similarly.

Notice the contrast with the elementarity in explicit mathematics: while the definedness predicate $\downarrow$ and the application $\circ$ can occur in elementary formulae in explicit mathematics, they cannot in Δ_0 formulae in weak explicit set theory. Nonetheless, by any axiom scheme defined for Δ_0 formulae we always mean the universal closures of the instances of the scheme and hence we can substitute any term t to the parameters in the Δ_0 formulae, as far as $t\downarrow$. Thus, loosely speaking, $\circ$ can occur but cannot apply to bounded variables.

Remark 16. $A \mapsto A^\approx$ is an embedding of $\mathcal{L}_\in$ into $\mathcal{L}_\in^{nso}$. By combining the standard embedding of the language $\mathcal{L}_2$ of second order arithmetic into $\mathcal{L}_\in$, we will consider $\mathcal{L}_2$ as a sublanguage of $\mathcal{L}_\in^{nso}$.

Furthermore we also use standard set-theoretic notations, but with $=$ replaced by $\approx$, as follows:

$$\begin{aligned}
a \subseteq b &:\equiv \mathcal{S}(a) \wedge \mathcal{S}(b) \wedge (\forall x \in a)(x \in b) \\
\mathrm{Trans}[a] &:\equiv \mathcal{S}(a) \wedge (\forall x \in a)(x \subseteq a) \\
\mathrm{Rel}[f] &:\equiv \mathcal{S}(f) \wedge (\forall z \in f)(\exists x, y)(z \approx \langle x, y\rangle) \\
\mathrm{Fun}[f] &:\equiv \mathrm{Rel}[f] \wedge (\forall x, y, z)(\langle x, y\rangle \in f \wedge \langle x, z\rangle \in f \ \rightarrow \ y \approx z) \\
\mathrm{Dom}[f, a] &:\equiv \mathrm{Rel}[f] \wedge (\forall x \in a)(\exists y)(\langle x, y\rangle \in f) \wedge (\forall x, y)(\langle x, y\rangle \in f \rightarrow x \in a) \\
\mathrm{Ran}[f, b] &:\equiv \mathrm{Rel}[f] \wedge (\forall x, y)(\langle x, y\rangle \in f \ \rightarrow \ y \in b).
\end{aligned}$$

Here, $\mathrm{Trans}[a]$ expresses that a is *transitive*, $\mathrm{Rel}[f]$ and $\mathrm{Fun}[f]$ express that f is a set-theoretic relation and function, respectively, $\mathrm{Dom}[f, a]$ expresses that the domain of f is a, and $\mathrm{Ran}[f, b]$ expresses that the range of f is a subset of b. By the usual proofs, we can show that these formulae are equivalent (over a very basic base theory) to some Δ_0 formulae, under the assumptions $\mathcal{S}(a), \mathcal{S}(b)$ and $\mathcal{S}(f)$.

Furthermore, we also introduce the following function-like abbreviations:

$$\begin{aligned}
f'x \approx y &:\equiv \mathrm{Fun}[f] \wedge \langle x, y\rangle \in f \\
a \approx \{b_0, ..., b_n\} &:\equiv \mathcal{S}(a) \wedge b_0 \in a \wedge ... \wedge b_n \in a \wedge (\forall x \in a)(x \approx b_0 \vee ... \vee x \approx b_n) \\
a \approx \bigcup b &:\equiv \mathcal{S}(a) \wedge (\forall x \in b)(x \subseteq a) \wedge (\forall x \in a)(\exists y \in b)(x \in y) \\
a \approx b \cup c &:\equiv \mathcal{S}(a) \wedge (\forall x \in a)(x \in b \vee x \in c) \wedge b \subseteq a \wedge c \subseteq a \\
a \approx b \setminus c &:\equiv \mathcal{S}(a) \wedge (\forall x \in a)(x \in b \wedge \neg x \in c) \wedge (\forall x \in b)(\neg x \in c \ \rightarrow \ x \in a) \\
a \approx \{x : \varphi(x)\} &:\equiv \mathcal{S}(a) \wedge (\forall x \in \mathcal{S})(x \in a \leftrightarrow \varphi(x)).
\end{aligned}$$

Note that, in the absence of extensionality, a in these abbreviations might not be unique even in the sense of the weaker equality $\approx$. Namely, $x \approx \{y, z\}$ and $x \approx \bigcup y$ mean "x is some set containing exactly y and z" and "x is some set corresponding to the union of y", respectively.

3.2. Weak explicit set theory with non-set operators

Definition 17. The theory WEST^{nso} is an $\mathcal{L}_{\in}^{nso}$-theory whose underlying logic is a classical logic of partial terms (due to Beeson [5]) with the equality axiom for $=$, and whose non-logical axioms are the following.

The axioms of Group A, the so-called *applicative axioms*, which are shared with explicit mathematics, are the standard axioms about the combinators and operations:

(A.1) $\mathbf{k}ab = a \ \wedge \ \mathbf{s}ab\downarrow \wedge \ \mathbf{s}abc \simeq (ac)(bc)$;

(A.2) $\mathbf{p}_0(\mathbf{p}(x, y)) = x \ \wedge \ \mathbf{p}_1(\mathbf{p}(x, y)) = y$.

The axioms of Group B are called *ontological axioms*, which govern the relation between non-sets and sets and the fundamental relation symbols $\mathcal{S}$, $\approx$ and $\in$:

(B.1) Set-hood $a \in b \vee a \approx b \ \rightarrow \ \mathcal{S}(a) \wedge \mathcal{S}(b)$;

(B.2) Equivalence Relation $a \approx a \wedge (a \approx b \rightarrow b \approx a) \wedge (a \approx b \wedge b \approx c \rightarrow a \approx c)$;

(B.3) Set Equality on Predicate $a \approx b \ \rightarrow \ \forall x((x \in a \leftrightarrow x \in b) \wedge (a \in x \leftrightarrow b \in x))$;

(B.4) Set Equality on Unary Operation $a \approx b \ \rightarrow \ \mathsf{c}(a) \approx \mathsf{c}(b)$, for $\mathsf{c} = \mathbb{K}, \mathbb{T}, \mathbb{U}, \mathbb{G}_1, \mathbb{G}_2$;

(B.5) Set Equality on Binary Operation $a_1 \approx b_1 \wedge a_2 \approx b_2 \ \rightarrow \ \mathsf{c}(a_1, a_2) \approx \mathsf{c}(b_1, b_2)$, for $\mathsf{c} = \mathbb{D}, \mathbb{G}_3, \mathbb{G}_4, \mathbb{G}_5, \mathbb{G}_{\approx}, \mathbb{G}_{\in}$;

(B.6) Set Equality on Replacement Operator $a \approx b \wedge (\forall x \in a)(fx \approx gx) \rightarrow \mathbb{R}(a, f) \approx \mathbb{R}(b, g)$;

(B.7) Foundation $(\forall x \in \mathcal{S})((\forall y \in x)A[y] \rightarrow A[x]) \ \rightarrow \ (\forall x \in \mathcal{S})A[x]$, for arbitrary $\mathcal{L}_{\in}^{nso}$-formula $A[x]$.

The axioms of Group C, the so-called *operational-set-theoretic axioms*, determine the roles of the symbols for the basic set-theoretic operations:

(C.1) Infinity $\mathcal{S}(\omega) \wedge \mathrm{Ind}[\omega] \wedge (\forall x \subseteq \omega)(\mathrm{Ind}[x] \to \omega \subseteq x)$,
where $\mathrm{Ind}[x]$ stands for $\emptyset \in x \wedge (\forall y \in x)(\exists z \in x)(z \approx y \cup \{y\})$ and $\emptyset$ stands for $\mathbb{G}_3(\omega, \omega)$;

(C.2) Kleene star $\mathcal{S}(a) \to \begin{array}{c} \mathbb{K}(a){\downarrow} \wedge (\exists u \in \mathbb{K}(a))(u \approx \emptyset) \wedge \\ (\forall u \in \mathbb{K}(a))(\exists n \in \omega) \begin{pmatrix} \mathrm{Fun}[u] \wedge \mathrm{Dom}[u, n] \wedge \mathrm{Ran}[u, a] \wedge \\ (\forall x \in a)(\exists v \in \mathbb{K}(a))(v \approx u \cup \{\langle n, x \rangle\}) \end{pmatrix} \end{array}$;

(C.3) Transitive closure $\mathcal{S}(a) \to \mathbb{T}(a){\downarrow} \wedge a \subseteq \mathbb{T}(a) \wedge \mathrm{Trans}[\mathbb{T}(a)] \wedge (\forall y)(x \subseteq y \wedge \mathrm{Trans}[y] \to \mathbb{T}(a) \subseteq y)$;

(C.4) Doubleton (or unordered pair) $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{D}(a, b){\downarrow} \wedge \mathbb{D}(a, b) \approx \{a, b\}$;

(C.5) Union $\mathcal{S}(a) \to \mathbb{U}(a){\downarrow} \wedge \mathbb{U}(a) \approx \bigcup a$;

(C.6) Replacement $\begin{array}{l} \mathcal{S}(a) \wedge (\forall x \in a)\mathcal{S}(fx) \wedge (\forall x, y \in a)(x \approx y \to fx \approx fy) \\ \to (\mathbb{R}(a, f){\downarrow} \wedge \forall x(x \in \mathbb{R}(a, f) \leftrightarrow (\exists y \in a)(x \approx fy))); \end{array}$

(C.7) Domain $\mathcal{S}(a) \to \mathbb{G}_1(a){\downarrow} \wedge \mathbb{G}_1(a) \approx \{v : (\exists w)(\langle v, w \rangle \in a)\}$;

(C.8) Range $\mathcal{S}(a) \to \mathbb{G}_2(a){\downarrow} \wedge \mathbb{G}_2(a) \approx \{w : (\exists v)(\langle v, w \rangle \in a)\}$;

(C.9) Difference $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{G}_3(a, b){\downarrow} \wedge \mathbb{G}_3(a, b) \approx a \setminus b$;

(C.10) Product $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{G}_4(a, b){\downarrow} \wedge \mathbb{G}_4(a, b) \approx \{\langle u, v \rangle : u \in a \wedge v \in b\}$;

(C.11) Permutation $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{G}_5(a, b){\downarrow} \wedge \mathbb{G}_5(a, b) \approx \{\langle u, \langle w, v \rangle \rangle : \langle u, v \rangle \in a \wedge w \in b\}$;

(C.12) Diagonalization $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{G}_{\approx}(a, b){\downarrow} \wedge \mathbb{G}_{\approx}(a, b) \approx \{\langle v, w \rangle \in a \times b : v \approx w\}$.

(C.13) Membership $\mathcal{S}(a) \wedge \mathcal{S}(b) \to \mathbb{G}_{\in}(a, b){\downarrow} \wedge \mathbb{G}_{\in}(a, b) \approx \{\langle v, w \rangle \in a \times b : v \in w\}$.

WEST_0^{nso} is the subsystem of WEST^{nso} with Foundation (B.7) removed completely; $\mathsf{WEST}_\omega^{nso}$ is the subsystem WEST^{nso} with Foundation relativized to ω (namely $\forall x \in \mathcal{S}$ replaced by $\forall x \in \omega$); WEST_r^{nso} is the subsystem WEST^{nso} with Foundation restricted to Δ_0 formulae; WEST_w^{nso} is $\mathsf{WEST}_r^{nso} + \mathsf{WEST}_\omega^{nso}$.

Note that the axiom of extensionality, even restricted to $\mathcal{S}$, is not included. All the operations mentioned in Group C assign sets to sets (except the first argument of $\mathbb{R}$). Nothing is assumed about the application of these operations to non-sets.

WEST^{int} introduced in Sato and Zumbrunnen [37, AppendixB] is essentially the same as $\mathsf{WEST}^{nso} + \forall x\mathcal{S}(x) + \forall x, y(x = y \leftrightarrow x \approx y) + \forall x(\forall y(y \in x \leftrightarrow y \approx a) \to \mathbb{U}(x) \approx a)$.

3.3. Basic properties

As in explicit mathematics, which shares the applicative axioms with weak explicit set theory, it is immediate to see the following.

Lemma 18. (i) For any $\mathcal{L}_{\in}^{nso}$-term $t[\vec{x}]$, there is an $\mathcal{L}_{\in}^{nso}$-term, denoted by $\lambda\vec{x}.t[\vec{x}]$, such that

$$\mathsf{WEST}_0^{nso} \vdash (\lambda\vec{x}.t[\vec{x}]){\downarrow} \wedge (\lambda\vec{x}.t[\vec{x}])\vec{y} \simeq t[\vec{y}].$$

(ii) There is an $\mathcal{L}_{\in}^{nso}$-term $\mathbf{fix}$ such that

$$\mathsf{WEST}_0^{nso} \vdash \mathbf{fix}(f){\downarrow} \wedge \mathbf{fix}(f)(x) \simeq f(\mathbf{fix}(f), x).$$

Following the standard argument on Gödel operations, we have the following.

Lemma 19. For any Δ_0 formula $A[y, \vec{x}]$ whose free variables are among $y, \vec{x}$, there is an $\mathcal{L}_\in^{nso}$-term s_A such that
$$\mathsf{WEST}_0^{nso} \vdash s_A(a, \vec{b}) \approx \{y \in a : A[y, \vec{b}]\}.$$

We have two notions of function: set-theoretic functions represented by graphs and operations represented by operators which form a non-total combinatory algebra with the application $\circ$. By the basically same proof as in Sato and Zumbrunnen [37], we can show the relation between these two notions as follows.

Proposition 20. There is a closed $\mathcal{L}_\in^{nso}$ term $\mathbb{F}$ such that WEST_0^{nso} proves

$$\mathcal{S}(a) \wedge (\forall x \in a)\mathcal{S}(fx) \wedge (\forall x, y \in a)(x \approx y \to fx \approx fy) \;\to\; \mathrm{Fun}[\mathbb{F}(f, a)] \wedge \mathrm{Dom}[\mathbb{F}(f, a), a] \wedge (\forall x \in a)(\mathbb{F}(f, a)' x \approx fx).$$

Proof. It is easy to see that $\mathbb{F} = \lambda f, a.\mathbb{R}(\lambda x.\langle x, fx\rangle, a)$ satisfies the stated property. □

Thus, loosely speaking, any extensional operation on a set is also a set-theoretic function. Sato and Zumbrunnen [37, Proposition 13] showed that the converse holds in WEST_0, namely there is a closed term **op** such that WEST_0 proves

$$\mathbf{op}(f)\downarrow \wedge ((\mathrm{Fun}[f] \wedge x \in \mathbb{G}_1(f)) \to f' x = \mathbf{op}(f, x)).$$

However, as pointed out there, we need the axiom of extensionality to prove this.

3.4. Axiom Beta

In the usual set-theoretic context, Axiom Beta asserts that any well-founded relation has a transitive (Mostowski) collapse. The operational version of this axiom was introduced in Sato and Zumbrunnen [37]. We modify it for the present context with $\approx$.

Definition 21. $\mathsf{WEST}^{nso}(\mathbb{B})$ is the extension of WEST^{nso} augmented by

$$\text{(C.14) Mostowski collapse} \left(\mathcal{S}(a) \wedge \mathrm{Rel}[r] \to \begin{array}{c} (a_1 \approx a_2 \wedge \mathrm{Rel}[r_1] \wedge \mathrm{Rel}[r_2] \wedge r_1 \approx r_2 \to \mathbb{B}(a_1, r_1) \approx \mathbb{B}(a_2, r_2)) \wedge \\ \mathrm{Fun}[\mathbb{B}(a, r)] \wedge \mathrm{DwCl}[\mathbb{G}_1(\mathbb{B}(a, r)), a, r] \wedge \\ \mathrm{Prog}[\mathbb{G}_1(\mathbb{B}(a, r)), a, r] \wedge \mathrm{Clp}[\mathbb{B}(a, r), \mathbb{G}_1(\mathbb{B}(a, r)), r] \end{array}\right),$$

where $\mathrm{DwCl}[b, a, r]$, $\mathrm{Prog}[b, a, r]$ and $\mathrm{Clp}[f, b, r]$ are the following formulae:

$$\begin{aligned} \mathrm{DwCl}[b, a, r] &:\equiv (\forall x \in b)(\forall y \in a)(\langle y, x\rangle \in r \to y \in b) \\ \mathrm{Prog}[b, a, r] &:\equiv (\forall x \in a)((\forall y \in a)(\langle y, x\rangle \in r \to y \in b) \to x \in b) \\ \mathrm{Clp}[f, b, r] &:\equiv (\forall x \in b)(f' x \approx \{f' y : y \in b \wedge \langle y, x\rangle \in r\}). \end{aligned}$$

$\mathsf{WEST}_0^{nso}(\mathbb{B})$, $\mathsf{WEST}_\omega^{nso}(\mathbb{B})$, $\mathsf{WEST}_r^{nso}(\mathbb{B})$ and $\mathsf{WEST}_w^{nso}(\mathbb{B})$ are defined similarly.

Intuitively, in the presence of Δ_0-foundation, $\mathbb{B}$ returns both the well-founded part of the given ordered structure (a, r) and the collapsing function of that part. However, the former is coded in the latter, as the former is the domain of the latter. Thus we design the operator to assign only the collapsing function of the well-founded part of the given ordered structure.

It was shown in Sato and Zumbrunnen [37, Lemma 31] that

$$\mathrm{Rel}[r] \;\to\; \exists f, b(\mathrm{Fun}[f] \wedge \mathrm{DwCl}[b, a, r] \wedge \mathrm{Prog}[b, a, r] \wedge \mathrm{Clp}[f, b, r])$$

is equivalent, in the presence of Δ_0-foundation, to Axiom Beta in the usual formulation:

$$(\forall b \subseteq a)(\mathrm{Prog}[b, a, r] \to a \subseteq b) \;\to\; \exists f(\mathrm{Fun}[f] \wedge \mathrm{Clp}[f, a, r]).$$

In this sense, (C.14) could be called the operational version of Axiom Beta. (See also the notion of operational Skolemization from Sato and Zumbrunnen [37, Subsection 7.4].)

4. Tree Interpretation of WEST^{nso} in Explicit Mathematics

In this section, we give an interpretation of weak explicit set theory with non-set operators in explicit mathematics. The basic idea is to represent the sets by trees formulated in explicit mathematics.

4.1. Tree of finite sequences of individuals

To realize the basic idea, we first prepare the notion of trees consisting of finite sequences of individuals within explicit mathematics, as in Figure 3 where $\lfloor$ is a closed $\mathbb{L}$-term and $s\lfloor\sigma$ denote $\lfloor(s,\sigma)$.

$$
\begin{aligned}
\mathrm{FSI} &:= \{\sigma : (\exists f, n)(n \in \mathsf{N} \wedge \sigma = \langle f, n\rangle \wedge (\forall k \in \mathsf{N})(fk\downarrow))\};\\
\mathrm{lh} &:= \mathbf{p}_1; \qquad \sigma[k] := \mathbf{p}_0(\sigma, k); \qquad \sigma\lceil n := \langle \mathbf{p}_0(\sigma), \min(n, \mathbf{p}_1(\sigma))\rangle;\\
\mathrm{shift}(\sigma) &:= \langle \lambda x.\mathbf{p}_0(\sigma, \mathsf{s}_\mathsf{N}(x)), \mathbf{p}_\mathsf{N}(\mathbf{p}_1(\sigma))\rangle;\\
\sigma * \tau &:= \langle \lambda x.\mathbf{d}_\mathsf{N}(\mathbf{p}_0(\sigma, x), \mathbf{p}_0(\tau, x \dot{-} \mathrm{lh}(\sigma)), \mathsf{s}_\mathsf{N}(x) \dot{-} \mathrm{lh}(\sigma), 0), \mathrm{lh}(\sigma) + \mathrm{lh}(\tau)\rangle;\\
\langle u\rangle &:= \langle \lambda x.u, 1\rangle; \qquad \langle u_0, ..., u_i\rangle * \sigma := \langle u_0\rangle * (... * (\langle u_i\rangle * \sigma)...);\\
\sigma \sim \tau &:\equiv \mathrm{lh}(\sigma) = \mathrm{lh}(\tau) \wedge (\forall k < \mathrm{lh}(\sigma))(\sigma[k] = \tau[k]);\\
\mathrm{Tree}[T] &:\equiv T \subseteq \mathrm{FSI} \wedge (\exists\sigma)(\sigma \in T) \wedge (\forall \sigma \in T)(\forall k < \mathrm{lh}(\sigma))(\sigma\lceil k \in T)\\
&\qquad \wedge (\forall \sigma, \tau \in \mathrm{FSI})(\sigma \sim \tau \to (\sigma \in T \leftrightarrow \tau \in T));\\
S \approx T &:\equiv (\forall \sigma \in \mathrm{FSI})(\sigma \in S \leftrightarrow \sigma \in T); \qquad s \approx t :\equiv (\exists S, T)(\Re(s, S) \wedge \Re(t, T) \wedge S \approx T);\\
&\Re(s\lfloor\sigma, \{v \in \mathrm{FSI} : \sigma * v \,\dot{\in}\, s\}).
\end{aligned}
$$

Figure 3: Notions for sequence and tree

Figure 3 introduces the notion of finite sequence of individuals and the notion of tree of such sequences, with some basic operations. FSI is the type of finite sequences of individuals; a sequence of individuals is a pair of an operation defined on all the natural numbers and a natural number called the length, where the values of the operation at numbers beyond the length have no meaning; $\mathrm{lh}(\sigma)$ denotes the length of the sequence σ; $\sigma[k]$ the k-th component of σ; $\sigma\lceil n$ the result of truncating σ at n; the operator shift drops the first component from a sequence; $*$ is the concatenation operator (in the definition of which, $\mathsf{p}_0(\sigma, x)$ must be defined for $x \geq \mathrm{lh}(\sigma)$ because of the strictness axiom of logic of partial terms); $\langle x_0, ..., x_k\rangle$ is the length k sequence whose i-th component is x_i (though this notation conflicts the previously introduced abbreviation $\langle s, t\rangle = \mathbf{p}(s, t)$, we employ the convention that, if the arguments are two, $\langle s, t\rangle$ always means $\mathbf{p}(s, t)$ and the other is denoted by $\langle s\rangle * \langle t\rangle$); and $\sigma \sim \tau$ means the two sequences σ and τ have the same length and exactly the same components. Furthermore, $\mathrm{Tree}[T]$ means that T is a tree on finite sequences of individuals, namely a non-empty type of finite sequences closed under truncations, where the identity on sequences are $\sim$ defined above. $S \approx T$ means that they contain exactly the same sequences. Though this is extensional from the viewpoint of the interpreting side, it is intensional from the interpreted side in the sense that, if $S \approx T$, there is no need to distinguish them. For this reason, we call such S and T isomorphic. Finally $s\lfloor\sigma \approx t\lfloor\tau$ means that the tree named by s truncated at σ is isomorphic to that by t at τ.

4.2. Tree representation of sets

Using the notions of finite sequence of individuals and of tree, Figure 4 defines the tree interpretations of important concepts of weak explicit set theory with non-set operators. In what follows, by $A^\natural[\vec{t}]$ we mean $(A[\vec{x}])^\natural[\vec{t}/\vec{x}]$. The so-called applicative part of weak explicit set theory is interpreted in the straightforward way, and therefore Group A of the axioms of WEST^{nso} are trivially interpreted in AET with this translation $\natural$. The sets are those trees, in the sense defined above, which are well-founded, in the sense of the special predicate $\Im$. The weaker equality $\approx$ is interpreted by the isomorphism defined previously, and therefore it is much stronger than the extensional equality or, equivalently, than the so-called bisimilarity, while it is weaker than the absolute equality $=$. The interpretation of the membership relation is natural from the

$$\mathbf{k}^\natural := \mathbf{k}; \quad \mathbf{s}^\natural := \mathbf{s}; \quad \mathbf{p}^\natural := \mathbf{p}; \quad (\mathbf{p}_0)^\natural := \mathbf{p}_0; \quad (\mathbf{p}_1)^\natural := \mathbf{p}_1; \quad x^\natural := x \text{ for any variable } x;$$

$$(t \circ s)^\natural := t^\natural \circ s^\natural; \quad (t\downarrow)^\natural :\equiv t^\natural\downarrow; \quad (s = t)^\natural :\equiv s^\natural = t^\natural;$$

$$(\mathcal{S}(t))^\natural :\equiv (\exists T)(\Re(t^\natural, T) \wedge \mathrm{Tree}[T] \wedge \Im(\{w : (\exists \sigma \in T)(w = \langle \sigma, \sigma\lceil(\mathrm{lh}(\sigma) - 1)\rangle \wedge 0 < \mathrm{lh}(\sigma))\}));$$

$$(s \approx t)^\natural :\equiv (\mathcal{S}(s))^\natural \wedge (\mathcal{S}(t))^\natural \wedge \{\tau \in \mathrm{FSI} : \tau \dot\in s^\natural\} \approx \{\tau \in \mathrm{FSI} : \tau \dot\in t^\natural\}$$

$$(s \in t)^\natural :\equiv (\mathcal{S}(s))^\natural \wedge (\mathcal{S}(t))^\natural \wedge (\exists x)(\{\tau \in \mathrm{FSI} : \tau \dot\in s^\natural\} \approx \{\tau \in \mathrm{FSI} : \langle x\rangle * \tau \dot\in t^\natural\});$$

$$\bot^\natural :\equiv \bot; \quad (A \wedge B)^\natural :\equiv A^\natural \wedge B^\natural; \quad (A \vee B)^\natural :\equiv A^\natural \vee B^\natural; \quad (A \to B)^\natural :\equiv A^\natural \to B^\natural;$$

$$(\forall x A[x])^\natural :\equiv \forall x(A[x]^\natural); \quad (\exists x A[x])^\natural :\equiv \exists x(A[x]^\natural).$$

Figure 4: Tree interpretations of basic notions of weak explicit set theory

standard tree representation of sets. Now it is easy to see that the first two (B.1) and (B.2) of Group B of the axioms of WEST^{nso} are interpreted in AET by $\natural$.

Also we can see that Foundation (B.7) is interpreted by $\natural$ in $\mathsf{AET} + (\mathbb{L}\text{-}\mathsf{TI})$ as follows. First we have to notice that $x \approx y$ implies $\mathcal{S}(x)^\natural \wedge A[x]^\natural \leftrightarrow \mathcal{S}(y)^\natural \wedge A[y]^\natural$. By definition, if $(\mathcal{S}(a))^\natural$ then $\Im(R)$ holds where $R = \{w : (\exists \sigma \dot\in a)(w = \langle \sigma, \sigma\lceil(\mathrm{lh}(\sigma) - 1)\rangle \wedge 0 < \mathrm{lh}(\sigma))\}$. To see that (B.7) is interpreted by $\natural$, assume $((\forall x \in \mathcal{S})((\forall y \in x)A[y] \to A[x]))^\natural$, i.e., $\forall x(\mathcal{S}(x)^\natural \wedge (\forall y)(y \in x \to A[y])^\natural \to A[x]^\natural)$. This implies

$$\forall \sigma \dot\in a(\forall u(\sigma * \langle u\rangle \dot\in a \to A^\natural[a\lfloor(\sigma * \langle u\rangle)]) \;\to\; A^\natural[a\lfloor\sigma]),$$

namely $\forall\sigma(\forall\tau(\langle\tau,\sigma\rangle \in R \to (\tau \dot\in a \to A^\natural[a\lfloor\tau])) \;\to\; (\sigma \dot\in a \to A^\natural[a\lfloor\sigma]))$. By transfinite induction along R, we have $\forall\sigma(\sigma \dot\in a \to A^\natural[a\lfloor\sigma])$ and, in particular, $A[a]^\natural$.

In order to turn to the other axioms, we first need to specify the interpretations of the constants for set generating operations. With the basic idea of tree representation in their minds, the readers can easily imagine how the interpretations should be like and convince themselves that it is possible. Figure 5 gives the explicit definitions of these interpretations for worried readers, but the author recommends the readers to try to define them in their favorite ways rather than to read the details of Figure 5. Note that, since $\{x\lfloor\langle u\rangle : \langle u\rangle \dot\in x\}$ covers $\{y : (y \in x)^\natural\}$ only up to $\approx$, the extensionality precondition for $\mathbb{R}$ is necessary.

The point of the figure is that, under the assumptions $\Re(x, X)$ and $\Re(y, Y)$, the defining formulae of the types occurring in $\Re$ as the second arguments are all elementary in X and Y, except that for $\mathbb{R}$. Thus uniform elementary comprehension (Lemma 6) gives the required interpretations as closed terms. For the definition of $\mathbb{R}$, however, we need the axiom for the join operator $\mathbf{j}$, as well as uniform elementary comprehension. It is obvious that these operations respect the isomorphism, namely the operations applied to isomorphic trees return isomorphic trees, and so we can see that the remaining axioms of Group B are interpreted by $\natural$, if we show that the values of these operations are in $\mathcal{S}$, which also implies the interpretability of Group C.

For ω, since $f = \lambda\sigma.\sigma[\mathrm{lh}(\sigma)\dot-1] := \lambda\sigma.\mathbf{p}_0(\sigma, \mathbf{p}_1(\sigma)\dot-1)$ satisfies $f\sigma < f(\sigma\lceil\mathrm{lh}(\sigma)\dot-1)$ for $\sigma \dot\in \omega^\natural$ with $1 < \mathrm{lh}(\sigma)$, the axioms (IV.2) and $(\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$ i.e., $\Im(<)$, imply $\Im(\{\langle\sigma, \sigma\lceil(\mathrm{lh}(\sigma)\dot-1)\rangle : \sigma \dot\in \omega^\natural, 0 < \mathrm{lh}(\sigma)\})$ and so $\mathcal{S}(\omega)^\natural$.

For $\mathbb{K}$, if $\mathcal{S}(x)^\natural$, either $\mathbb{K}^\natural(x)\lfloor(\sigma\lceil 5) \approx \omega^\natural\lfloor(\mathbf{p}_0(\sigma[1]))$ or $\mathbb{K}^\natural(x)\lfloor(\sigma\lceil 5) \approx x\lfloor(\mathbf{p}_1(\sigma[1]))$ for any $\sigma \dot\in \mathbb{K}^\natural(x)$ with $\mathrm{lh}(\sigma) \geq 5$, and so the axiom (IV.2) implies $\mathcal{S}^\natural(\mathbb{K}(x)^\natural\lfloor(\sigma\lceil 5))$. By applying (VI.3) step-by-step, we have $\mathcal{S}(\mathbb{K}(x)^\natural\lfloor(\sigma\lceil 4))$, $\mathcal{S}^\natural(\mathbb{K}^\natural(x)\lfloor(\sigma\lceil 3))$, $\mathcal{S}^\natural(\mathbb{K}^\natural(x)\lfloor(\sigma\lceil 2))$, $\mathcal{S}^\natural(\mathbb{K}^\natural(x)\lfloor(\sigma\lceil 1))$ and then finally $\mathcal{S}(\mathbb{K}(x))^\natural$.

We can treat the other set generators similarly.

Moreover, for any name $a \dot\subseteq \mathsf{nat}$, the corresponding subset $\{\sigma \dot\in \omega^\natural : \mathrm{lh}(\sigma) > 0 \to \sigma[0] \dot\in a\}$ can be constructed uniformly, and conversely if $(a \subseteq \omega)^\natural$ then $\{n \in \mathsf{N} : \langle n\rangle \dot\in a\}$ forms a type. Thus the interpretation preserves all the formulae from $\mathcal{L}_2$ up to the equivalence.

Theorem 22. WEST^{nso} is interpreted by $\natural$ in $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N}) + (\mathbb{L}\text{-}\mathsf{TI})$ in an $\mathcal{L}_2$-preserving way.

Moreover, under the assumptions $(\mathcal{S}(s))^\natural$, $(\mathcal{S}(t))^\natural$, $\Re(s^\natural, S)$ and $\Re(t^\natural, T)$, the formula $(s \in t)^\natural$ is equivalent to an elementary formula in S and T. Therefore in order to interpret foundation restricted to Δ_0 we need only a type transfinite induction, which is included in AETJ, as seen below.

$$\Re(\omega^\natural,\ \{\sigma \in \mathrm{FSI} : (\forall k \in \mathsf{N})(k < \mathrm{lh}(\sigma) \to \sigma[k] \in \mathsf{N}) \wedge (\forall k, l \in \mathsf{N})(k < l < \mathrm{lh}(\sigma) \to \sigma[k] > \sigma[l])\});$$

$$\Re\left(\mathbb{K}^\natural(x),\ \left\{\sigma \in \mathrm{FSI} : \begin{array}{c} (\mathrm{lh}(\sigma) > 0 \to (\exists n \in \mathsf{N})(\exists f)(\sigma[0] = \langle n, f\rangle \wedge (\forall k < n)(\langle fk\rangle \mathrel{\dot\in} x))) \\ (\mathrm{lh}(\sigma) > 1 \to (\exists k \in \mathsf{N})(k < \mathbf{p_0}(\sigma[0]) \wedge \sigma[1] = \langle k, \mathbf{p_1}(\sigma[0])(k)\rangle)) \\ \wedge(\mathrm{lh}(\sigma) > 2 \to \sigma[2] = 0 \vee \sigma[2] = 1) \wedge (\mathrm{lh}(\sigma) > 3 \to \sigma[3] = 0 \vee \sigma[3] = 1) \\ \wedge (\mathrm{lh}(\sigma) > 4 \wedge (\sigma[2] = 0 \vee \sigma[3] = 0) \to \sigma[4] = \mathbf{p_0}(\sigma[1]) \wedge \mathrm{shift}^4(\sigma) \mathrel{\dot\in} \omega^\natural) \\ \wedge (\mathrm{lh}(\sigma) > 4 \wedge (\sigma[2] = 1 \wedge \sigma[3] = 1) \to \sigma[4] = \mathbf{p_1}(\sigma[1]) \wedge \mathrm{shift}^4(\sigma) \mathrel{\dot\in} x) \end{array}\right\}\right);$$

$$\Re(\mathbb{T}^\natural(x),\ \{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \sigma[0] \in \mathrm{FSI} \wedge \sigma[0] * \mathrm{shift}(\sigma) \mathrel{\dot\in} x\});$$

$$\Re(\mathbb{D}^\natural(x, y),\ \{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to (\sigma[0] = 0 \wedge \mathrm{shift}(\sigma) \mathrel{\dot\in} x) \vee (\sigma[0] = 1 \wedge \mathrm{shift}(\sigma) \mathrel{\dot\in} y)\});$$

$$\Re(\mathbb{U}^\natural(x),\ \{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \langle \mathbf{p_0}(\sigma[0])\rangle * \langle \mathbf{p_1}(\sigma[0])\rangle * \mathrm{shift}(\sigma) \mathrel{\dot\in} x\});$$

$$\Re(\mathbb{R}^\natural(f, x),\ \{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \langle \sigma[0], \mathrm{shift}(\sigma)\rangle \mathrel{\dot\in} \mathbf{j}(\{u \mathrel{\dot\in} x : \mathrm{lh}(u) = 1\}, \lambda u.f(x\lfloor u))\});$$

$$\Re\left((\mathbb{G}_1)^\natural(x),\ \left\{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \begin{pmatrix} \langle \sigma[0], 0, 0\rangle * \mathrm{shift}(\sigma) \mathrel{\dot\in} x \wedge \langle \sigma[0], 1, 1\rangle \mathrel{\dot\in} x \\ \wedge(x\lfloor\langle \sigma[0], 0, 0\rangle \approx x\lfloor\langle \sigma[0], 0, 1\rangle \approx x\lfloor\langle \sigma[0], 1, 0\rangle) \end{pmatrix}\right\}\right);$$

$$\Re\left((\mathbb{G}_2)^\natural(x),\ \left\{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \begin{pmatrix} \langle \sigma[0], 1, 1\rangle * \mathrm{shift}(\sigma) \mathrel{\dot\in} x \wedge \langle \sigma[0], 0, 0\rangle \mathrel{\dot\in} x \\ \wedge(x\lfloor\langle \sigma[0], 0, 0\rangle \approx x\lfloor\langle \sigma[0], 0, 1\rangle \approx x\lfloor\langle \sigma[0], 1, 0\rangle) \end{pmatrix}\right\}\right);$$

$$\Re((\mathbb{G}_3)^\natural(x, y),\ \{\sigma \in \mathrm{FSI} : \mathrm{lh}(\sigma) > 0 \to \sigma \mathrel{\dot\in} x \wedge \neg(\exists u)(x\lfloor\langle\sigma[0]\rangle \approx y\lfloor\langle u\rangle)\});$$

$$\Re\left((\mathbb{G}_4)^\natural(x, y),\ \left\{\sigma \in \mathrm{FSI} : \begin{array}{c} (\mathrm{lh}(\sigma) > 0 \to (\exists u, v)(\sigma[0] = \langle u, v\rangle \wedge \langle u\rangle \mathrel{\dot\in} x \wedge \langle v\rangle \mathrel{\dot\in} y)) \\ \wedge(\mathrm{lh}(\sigma) > 1 \to \sigma[1] = 0 \vee \sigma[1] = 1) \wedge (\mathrm{lh}(\sigma) > 2 \to \sigma[2] = 0 \vee \sigma[2] = 1) \\ \wedge(\mathrm{lh}(\sigma) > 2 \wedge (\sigma[2] = 0 \vee \sigma[1] = 0) \to \langle \mathbf{p_0}(\sigma[0])\rangle * \mathrm{shift}^3(\sigma) \mathrel{\dot\in} x) \\ \wedge(\mathrm{lh}(\sigma) > 2 \wedge (\sigma[2] = 1 \wedge \sigma[1] = 1) \to \langle \mathbf{p_1}(\sigma[0])\rangle * \mathrm{shift}^3(\sigma) \mathrel{\dot\in} y) \end{array}\right\}\right);$$

$$\Re\left((\mathbb{G}_5)^\natural(x, y),\ \left\{\sigma \in \mathrm{FSI} : \begin{array}{c} (\mathrm{lh}(\sigma) > 0 \to \begin{pmatrix} \langle \mathbf{p_0}(\sigma[0]), 0, 0\rangle \mathrel{\dot\in} x \wedge \langle \mathbf{p_0}(\sigma[0]), 1, 1\rangle \mathrel{\dot\in} x \wedge \langle \mathbf{p_1}(\sigma[0])\rangle \mathrel{\dot\in} y \\ (x\lfloor\langle\langle \mathbf{p_0}(\sigma[0]), 0, 0\rangle \approx x\lfloor\langle \mathbf{p_0}(\sigma[0]), 1, 0\rangle \approx x\lfloor\langle \mathbf{p_0}(\sigma[0]), 0, 1\rangle)) \end{pmatrix} \\ \wedge(\mathrm{lh}(\sigma) > 1 \to \sigma[1] = 0 \vee \sigma[1] = 1) \wedge (\mathrm{lh}(\sigma) > 2 \to \sigma[2] = 0 \vee \sigma[2] = 1) \\ \wedge \big(\mathrm{lh}(\sigma) > 3 \wedge (\sigma[1] = 0 \vee \sigma[1] = 0) \to \langle \mathbf{p_0}(\sigma[0]), 0, 0\rangle * \mathrm{shift}^3(\sigma) \mathrel{\dot\in} x\big) \\ \wedge(\mathrm{lh}(\sigma) > 3 \wedge (\sigma[1] = \sigma[2] = 1) \to \sigma[3] = 0 \vee \sigma[3] = 1) \\ \wedge(\mathrm{lh}(\sigma) > 4 \wedge (\sigma[1] = \sigma[2] = 1) \to \sigma[4] = 0 \vee \sigma[4] = 1) \\ \wedge \begin{pmatrix} \mathrm{lh}(\sigma) > 5 \wedge (\sigma[1] = \sigma[2] = 1) \wedge (\sigma[3] = 0 \vee \sigma[4] = 0) \\ \to \langle \mathbf{p_1}(\sigma[0])\rangle * \mathrm{shift}^5(\sigma) \mathrel{\dot\in} y \end{pmatrix} \\ \wedge \begin{pmatrix} \mathrm{lh}(\sigma) > 5 \wedge (\sigma[1] = \sigma[2] = \sigma[3] = \sigma[4] = 1) \\ \to \langle \mathbf{p_0}(\sigma[0]), 1, 1\rangle * \mathrm{shift}^5(\sigma) \mathrel{\dot\in} x \end{pmatrix} \end{array}\right\}\right);$$

$$\Re\left((\mathbb{G}_\approx)^\natural(x, y),\ \left\{\sigma \mathrel{\dot\in} (\mathbb{G}_4)^\natural(x, y) : \mathrm{lh}(\sigma) > 0 \to x\lfloor\langle \mathbf{p_0}(\sigma[0])\rangle \approx y\lfloor\langle \mathbf{p_1}(\sigma[0])\rangle\right\}\right);$$

$$\Re\left((\mathbb{G}_\in)^\natural(x, y),\ \left\{\sigma \mathrel{\dot\in} (\mathbb{G}_4)^\natural(x, y) : \mathrm{lh}(\sigma) > 0 \to (\exists z)\begin{pmatrix} \langle \mathbf{p_1}(\sigma[0])\rangle * \langle z\rangle \mathrel{\dot\in} y \wedge \\ x\lfloor\langle \mathbf{p_0}(\sigma[0])\rangle \approx y\lfloor(\langle \mathbf{p_1}(\sigma[0])\rangle * \langle z\rangle) \end{pmatrix}\right\}\right).$$

Figure 5: Definitions of tree interpretations of set generators

Theorem 23. WEST^{nso}_r and WEST^{nso}_w are interpreted by $\natural$ in $\mathsf{AETJ}+(\mathsf{T}\text{-}\mathsf{I_N})$ and $\mathsf{AETJ}+(\mathbb{L}\text{-}\mathsf{I_N})$ respectively, in an $\mathcal{L}_2$-preserving way.

Proof. Let $A[x]$ be a Δ_0 formula. Assume $((\forall x \in \mathcal{S})((\forall y \in x)A[y] \to A[x]))^\natural$. To show $((\forall x \in \mathcal{S})A[x])^\natural$, fix a with $\mathcal{S}(a)^\natural$. It suffices to show $A[a]^\natural$. We can take b such that $(a \in b)^\natural$ and $\mathrm{Trans}[b]^\natural$. Let s_A be the $\mathcal{L}^{nso}_\in$-term from Lemma 19, which means $(s_A(b) \approx \{y \in b : A[y]\})^\natural$. Then $(\forall x \in b)((\forall y \in x)(y \in s_A(b)) \to x \in s_A(b))^\natural$. As mentioned above, under the assumptions $(\mathcal{S}(s))^\natural$, $(\mathcal{S}(t))^\natural$, $\Re(s^\natural, S)$ and $\Re(t^\natural, T)$, the formula $(s \in t)^\natural$ is equivalent to an elementary formula in S and T and hence the proof previously given for full foundation scheme now works with type induction. $\square$

4.3. Interpreting Axiom Beta

With the inductive generation operator $\mathbf{ig}$, we can interpret (C.14) by defining $\mathbb{B}^\natural$ as follows, where we use auxiliary terms $\mathbf{a}$ and $\mathbf{r}$: $\mathbf{a}(x)$ names the type of all such u that the u-th immediate subtree of x is non-empty (i.e., the position u codes an element); and $\mathbf{r}(x,y)$ names the type of all the pairs $\langle u,v\rangle$ such that the u-th immediate subtree and v-th one are related by y. Therefore $\mathbf{ig}(\mathbf{a}(x),\mathbf{r}(x,y))$ names the type of all such u's that the u-th immediate subtree represents the set in the domain of the collapsing function represented by $\mathbb{B}^\natural(x,y)$, and hence we have $\Re((\mathbb{G}_1)^\natural(\mathbb{B}^\natural(x,y)),\{\sigma\,\dot{\in}\,x : \mathrm{lh}(\sigma)>0\to\sigma[0]\,\dot{\in}\,\mathbf{ig}(\mathbf{a}(x),\mathbf{r}(x,y))\})$.

$$\Re(\mathbf{a}(x),\ \{u : \langle u\rangle\,\dot{\in}\,x\});$$
$$\Re(\mathbf{r}(x,y),\ \{w : \langle w\rangle\,\dot{\in}\,\mathbb{G}_4(x,x)\ \wedge\ (\exists\sigma\,\dot{\in}\,y)(\mathrm{lh}(\sigma)=1\ \wedge\ \mathbb{G}_4(x,x)\lfloor\langle w\rangle\approx y\lfloor\sigma)\};$$
$$\Re\left(\mathbb{B}^\natural(x,y),\ \left\{\sigma\in\mathrm{FSI} : \begin{array}{c}(\mathrm{lh}(\sigma)>0\ \to\ \sigma[0]\,\dot{\in}\,\mathbf{ig}(\mathbf{a}(x),\mathbf{r}(x,y)))\\ \wedge(\mathrm{lh}(\sigma)>1\ \to\ \sigma[1]=0\vee\sigma[1]=1)\wedge(\mathrm{lh}(\sigma)>2\ \to\ \sigma[2]=0\vee\sigma[2]=1)\\ \wedge(\mathrm{lh}(\sigma)>3\wedge(\sigma[1]=0\vee\sigma[2]=0)\ \to\ \langle\sigma[0]\rangle*\mathrm{shift}^3(\sigma)\,\dot{\in}\,x)\\ \wedge\left(\begin{pmatrix}\mathrm{lh}(\sigma)>3\,\wedge\\ (\sigma[1]=\sigma[2]=1\end{pmatrix}\to\begin{pmatrix}\langle\sigma[3],\sigma[0]\rangle\,\dot{\in}\,\mathbf{r}(x,y)\,\wedge\\ (\forall k\in\mathsf{N})(2<k<\mathrm{lh}(\sigma)\to\sigma[k]\,\dot{\in}\,\mathbf{a}(x))\,\wedge\\ (\forall k\in\mathsf{N})(k+4<\mathrm{lh}(\sigma)\to\langle\sigma[k+4],\sigma[k+3]\rangle\,\dot{\in}\,\mathbf{r}(x,y))\end{pmatrix}\right)\end{array}\right\}\right).$$

We need to show $\mathcal{S}(x)^\natural\wedge\mathcal{S}(y)^\natural\ \to\ \mathcal{S}(\mathbb{B}(x,y))^\natural$. We write $\mathcal{S}^\natural(x)$ for $\Im(\{\langle\sigma,\sigma\restriction(\mathrm{lh}(\sigma)\dot{-}1)\rangle : \sigma\,\dot{\in}\,x, 0<\mathrm{lh}(\sigma)\})$. Assume $\mathcal{S}(x)^\natural$ and $\mathcal{S}(y)^\natural$. For $\sigma\,\dot{\in}\,\mathbb{B}^\natural(x,y)$, if $\mathrm{lh}(\sigma)>2$ and $\sigma[1]=0\vee\sigma[2]=0$ then $\mathbb{B}^\natural(x,y)\lfloor(\sigma\restriction 3)\approx x\lfloor\langle\sigma[0]\rangle$ and so $\mathcal{S}^\natural(\mathbb{B}^\natural(x,y)\lfloor(\sigma\restriction 3))$. By the axiom (III.8), $\Im(\{\langle u,v\rangle\,\dot{\in}\,\mathbf{r}(x,y) : v\,\dot{\in}\,\mathbf{ig}(\mathbf{a}(x),\mathbf{r}(x,y))\})$ and so $\lambda\sigma.\sigma[\mathrm{lh}(\sigma)\dot{-}1]$, with (II.2), witnesses $\mathcal{S}^\natural(\mathbb{B}^\natural(x,y)\lfloor(\sigma\restriction 3))$ for $\sigma\,\dot{\in}\,\mathbb{B}^\natural(x,y)$ with $\mathrm{lh}(\sigma)>2$ and $\sigma[1]=\sigma[2]=1$. By applying (VI.3) step-by-step, we have $\mathcal{S}(\mathbb{B}^\natural(x,y)\lfloor(\sigma\restriction 2))$, $\mathcal{S}^\natural(\mathbb{B}^\natural(x,y)\lfloor(\sigma\restriction 1))$ and then finally $\mathcal{S}^\natural(\mathbb{B}^\natural(x,y))$, i.e., $\mathcal{S}(\mathbb{B}(x,y))^\natural$.

Theorem 24. $\mathsf{WEST}^{nso}(\mathbb{B})$ is interpreted by $\natural$ in T_0 in an $\mathcal{L}_2$-preserving way. Similarly $\mathsf{WEST}^{nso}_r(\mathbb{B})$ and $\mathsf{WEST}^{nso}_w(\mathbb{B})$ are interpreted by $\natural$ in $\mathsf{T}_0\!\restriction$ and $\mathsf{T}_0\!\restriction+(\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ respectively, in an $\mathcal{L}_2$-preserving way.

4.4. Dropping foundation

The reason why we define $\mathcal{S}(t)^\natural$ with the predicate $\Im$ in Figure 4 is that we want to have foundation scheme interpreted by $\natural$. If we do not want it, we can drop this clause, namely, we can define $\sharp$ in the same way as $\natural$ except:

- $\mathcal{S}(t)^\sharp\ :\equiv\ \exists T(\Re(t^\sharp,T)\wedge\mathrm{Tree}[T])$;
- in the definition of $\mathbb{B}^\sharp$, the first line is $\mathrm{lh}(\sigma)>0\ \to\ \sigma[0]\,\dot{\in}\,\mathbf{a}(x)$.

With this modification, still we have $\mathcal{S}^\sharp(\omega^\natural)$ and so the induction schemata on natural numbers are interpreted by $\sharp$ in $\mathsf{AET}+(\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$ or $\mathsf{AET}+(\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$. The other axioms, except foundation schemata, are interpreted by $\sharp$, as we can show in the same way as in the case of $\natural$. Thus we have the following results:

Theorem 25. $\mathsf{WEST}^{nso}_0(\mathbb{B})$ and $\mathsf{WEST}^{nso}_\omega(\mathbb{B})$ are interpreted by $\sharp$ in $\mathsf{AETJ}+(\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$ and $\mathsf{AETJ}+(\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ respectively, in such a way that all $\mathcal{L}_2$-formulae are preserved.

Remark 26. Actually, in the theorem, the axiom for $\mathbb{B}$ can be strengthened as follows:

$$\text{(C.15) Anti-regularity}\ \begin{array}{c}(a_1\approx a_2\wedge\mathrm{Rel}[r_1]\ \wedge\ \mathrm{Rel}[r_2]\ \wedge\ r_1\approx r_2\ \to\ \mathbb{B}(a_1,r_1)\approx\mathbb{B}(a_2,r_2))\\ \wedge(\mathcal{S}(a)\ \wedge\ \mathrm{Rel}[r]\to\mathrm{Fun}[\mathbb{B}(a,r)]\ \wedge\ \mathrm{Clp}[\mathbb{B}(a,r),a,r])\end{array}.$$

This can be seen as an operational version of AFA_1 from Aczel [2], the statement that any relation has a transitive collapse. Aczel's famous anti-foundation axiom is the conjunction of AFA_1 and AFA_2, the latter of which states the uniqueness of the transitive collapse. The tree representation shows that, in the absence of foundation, AFA_1 is rather weak and AFA_2 is generally strong. Recall that, without the axiom of extensionality, we cannot prove the uniqueness of the transitive collapses of even well-founded relations. The uniqueness of the collapses of non-well-founded relations is a strong version of extensionality. Actually, in the context of Kripke-Platek set theory, Rathjen [34] shows that KP_ω with anti-foundation axiom has the same proof-theoretic strength as Σ^1_2-AC, and is much stronger than KP_ω which has the same as Σ^1_1-AC.

4.5. Interpretability in applicative theories with type-2 functionals but without types

Actually, the typing machinery is not essential for the interpretations $\natural$ and $\sharp$. $\mathsf{BON}(\boldsymbol{\mu})$, the basic theory of operations and numbers with non-constructive μ-operator $\boldsymbol{\mu}$, from Feferman and Jäger [13], has the applicative axioms and the axiom for non-constructive μ-operator

$$(\forall x \in \mathsf{N})(fx \in \mathsf{N}) \rightarrow (\boldsymbol{\mu}(f) \in \mathsf{N} \wedge (\forall x \in \mathsf{N})(fx = 0 \rightarrow f(\boldsymbol{\mu}(f)) = 0)).$$

$\mathsf{BON}(\boldsymbol{\mu})$ with basic or full induction can interpret ACA_0 or ACA respectively, by interpreting sets as operators on N with values 0 and 1 only (i.e., $(\forall x \in \mathsf{N})(fx = 0 \vee fx = 1)$). Via this interpretation, $\mathsf{BON}(\boldsymbol{\mu})$ can encode the basic notions for trees on natural numbers. Thus, by setting the interpretation of the predicate $\mathcal{S}$ as being a tree in this sense, we can define an interpretation similar to $\sharp$ so that the codes of operations corresponding to our set generators $\mathbb{K}$, $\mathbb{T}$, ... are definable by operators. Particularly, the definability of the replacement operator $\mathbb{R}$ is by the applicative nature (that is, not by any additional operation like join $\mathbf{j}$ in explicit mathematics) as follows: if f is an operator which assigns (codes of) sets to (codes of) sets,

$$\mathbb{R}(f, a)(\sigma) := \begin{cases} 1 & \text{if } \mathrm{lh}(\sigma) = 0 \\ f(a\lfloor\langle\sigma[0]\rangle)(\mathrm{shift}(\sigma)) & \text{if } \mathrm{lh}(\sigma) \geq 1 \end{cases}$$

can be defined by the applicative structure, where $a\lfloor\tau = \lambda\sigma.a(\tau * \sigma)$ codes the tree a truncated by τ. Here the case distinction is possible because it is by a primitive recursive predicate. Thus we can interpret

- WEST_0^{nso} in $\mathsf{BON}(\boldsymbol{\mu})$ plus basic induction on numbers;
- $\mathsf{WEST}_\omega^{nso}$ in $\mathsf{BON}(\boldsymbol{\mu})$ plus full induction on numbers.

Moreover, if we add a new predicate $\mathbf{Wf}$ and relevant axioms (like (IV.1-3) for $\Im$), and if we restrict the interpretation of $\mathcal{S}$ to those trees which are well-founded in the sense of $\mathbf{Wf}$, we can, similarly to $\natural$, interpret

- WEST_r^{nso} in $\mathsf{BON}(\boldsymbol{\mu})$ plus basic transfinite induction along $\mathbf{Wf}$-well-founded trees;
- WEST_w^{nso} in $\mathsf{BON}(\boldsymbol{\mu})$ plus full induction on numbers and basic transfinite induction along $\mathbf{Wf}$-well-founded trees;
- WEST^{nso} in $\mathsf{BON}(\boldsymbol{\mu})$ plus full transfinite induction along $\mathbf{Wf}$-well-founded trees.

Similarly to $\Im$ in explicit mathematics, in the first two cases, $\mathbf{Wf}$ can be taken as the defined well-foundedness from second order arithmetic, whereas in the last case it is intended to denote the "real" well-foundedness stronger than the defined well-foundedness.

In Jäger and Strahm [22] and Jäger and Probst [20], the theory is extended further by the so-called Suslin operator $\boldsymbol{S}$, which checks whether the codes of trees encode well-founded ones or not by returning 0 or 1. With this operation, the two notions of well-foundedness, one in the sense of the predicate $\mathbf{Wf}$ and the defined one, are equivalent as in explicit mathematics, and so there is no need to add the predicate. With $\boldsymbol{S}$ we can define the well-founded part of a given relation, and therefore we can interpret

- $\mathsf{WEST}_r^{nso}(\mathbb{B})$ in $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus basic induction on numbers;
- $\mathsf{WEST}_w^{nso}(\mathbb{B})$ in $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus full induction on numbers;
- $\mathsf{WEST}^{nso}(\mathbb{B})$ in $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus full bar induction.

Thus, our lower bound results that we will establish also yield the lower bounds for these applicative theories with type-2 functionals. While the lower bounds for those without bar induction were already given in the previous works (Feferman and Jäger [13], Jäger and Strahm [22] and Jäger and Probst [20]), the lower bounds for those with bar induction might be new, even in the sense of proof-theoretic strengths. Namely,

1. KP is interpretable in $\mathsf{BON}(\boldsymbol{\mu})$ plus full transfinite induction along $\mathbf{Wf}$-well-founded trees;
2. Both $\Sigma_2^1\text{-}\mathsf{AC} + (\mathsf{BI})$ and $\mathsf{KP}\beta$ are interpretable in $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus full bar induction.

The upper bounds of these variants of $\mathsf{BON}(\boldsymbol{\mu})$ and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ can be established by the so-called inductive model construction, within KP (as Feferman and Jäger [13]) and KPi (as Jäger and Strahm [22]) respectively, where the interpretation of KPi in $\Sigma_2^1\text{-}\mathsf{AC} + (\mathsf{BI})$ was given in Jäger [17, Section 8].

5. Realizability Interpretation

Sato and Zumbrunnen [37] embedded a version of intuitionistic Kripke-Platek set theory IKP^- into the weak explicit set theory WEST by a realizability interpretation. This interpretation can work for our weaker WEST^{nso} as well, with some appropriate modifications, as we will see in this section.

5.1. The intuitionistic theory IKP^-

Definition 27 (Negative and strongly negative formulae)**.** An $\mathcal{L}_\in$ formula is called *negative* if it is built up from atomic formulae by means of the connectives $\wedge$ and $\to$ and the quantifier $\forall$.

The *strongly negative* formulae are inductively defined as follows:

1. $\bot$ is a strongly negative formula;
2. if A is atomic and B strongly negative, then also $A \to B$ is strongly negative;
3. if both A and B are strongly negative, then so are $A \to B$, $A \wedge B$ as well as $\forall x A$.

Definition 28 (Gödel-Gentzen negative interpretation)**.** The *negative interpretation* A^N of each $\mathcal{L}_\in$ formula A is inductively defined as follows:

$$A^N :\equiv \neg\neg A \text{ if } A \text{ is atomic} \quad (B \wedge C)^N :\equiv B^N \wedge C^N \qquad (B \to C)^N :\equiv B^N \to C^N \quad (\forall x B)^N :\equiv \forall x(B^N)$$
$$(B \vee C)^N :\equiv \neg(\neg B^N \wedge \neg C^N) \qquad (\exists x B)^N :\equiv \neg\forall x \neg(B^N).$$

By definition, A^N is strongly negative and is classically equivalent to A. It is well known that if A is classically valid then A^N is intuitionistically valid.

Definition 29. The system IKP^-, formulated in the language $\mathcal{L}_\in$, is based on intuitionistic logic with equality axioms and consists of the following non-logical axioms:

(IKP^-.1) transitive superset $\exists x(a \subseteq x \wedge \mathrm{Trans}[x])$;

(IKP^-.2) (weak) pairing $\exists x(a \in x \wedge b \in x)$;

(IKP^-.3) (weak) union $\exists x(\forall y \in a)(\forall z \in y)(z \in x)$;

(IKP^-.4) Δ_0^- separation $\exists x((\forall y \in x)(y \in a \wedge A[y]) \wedge (\forall y \in a)(A[y] \to y \in x))$
for all negative Δ_0 formulae $A[y]$ in which x does not occur;

(IKP^-.5) Δ_0 collection $(\forall x \in a)\exists y A[x,y] \to \exists z(\forall x \in a)(\exists y \in z)A[x,y]$
for all Δ_0 formulae $A[x,y]$ in which z does not occur;

(IKP^-.6) $\in$-induction for all $\mathcal{L}_\in$ formulae $\forall x((\forall y \in x)A[y] \to A[x]) \to \forall x A[x]$ for all formulae $A[x]$

(IKP^-.7) $\mathrm{Ind}[\omega]^N \wedge (\forall x \subseteq \omega)(\mathrm{Ind}[x] \to \omega \subseteq x)^N$ (N-infinity),
where $\mathrm{Ind}[x] \equiv ((\exists y \in x)\mathrm{zero}[y] \wedge (\forall y \in x)(\exists z \in x)\mathrm{isucc}[y,z])$, $\mathrm{zero}[y]$ is the formula $(\forall z \in y)\bot$, $\mathrm{isucc}[y,z]$ is the conjunction of the three formulae $y \in z$, $(\forall u \in y)(u \in z)$ and $(\forall u \in z)(u \in y \vee u = y)$;

(IKP^-.8) $(\exists z)\mathrm{famfun}[\omega,y,z]^N$ (N-Kleene star)
where $\mathrm{famfun}[\omega,y,z]$ is the conjunction of "z is a set of functions from some natural numbers to y" $(\forall u \in z)(\exists n \in \omega)(\mathrm{Fun}[u] \wedge \mathrm{Dom}[u,n] \wedge \mathrm{Ran}[u,y])$, of "$z$ contains an empty sequence" $(\exists u \in z)(\forall x \in u)\bot$, and of "any sequence in z can be extended by putting any element of y at the end" $(\forall u \in z)(\forall n \in \omega)(\mathrm{Dom}[u,n] \to (\forall x \in y)(\exists v \in z)(\mathrm{Dom}[v,n+1] \wedge u \subseteq v \wedge \langle n,x\rangle \in v))$.

IKP_0^- is IKP^- minus $\in$-induction; IKP_ω^- is IKP_0^- plus $\in$-induction on ω for all $\mathcal{L}_\in$ formulae; IKP_r^- is IKP_0^- plus $\in$-induction for all negative Δ_0 formulae; and IKP_w^- is $\mathsf{IKP}_\omega^- + \mathsf{IKP}_r^-$.

In order to derive non-weak pairing and union (i.e., those in the usual formulations) from weak ones (as above), it seems necessary to use the separation for non-negative Δ_0 formulae (cf. [37, p.147, the last three lines]).

These theories are in a sense constructive, because they are subsystems of constructive Zermelo-Fraenkel set theory CZF (for the constructive justification of CZF, see Aczel [1]). Also, the theories are intensional in the sense that they do not include the axiom of extensionality.

5.2. Extensional realizability notion

As opposed to WEST^{int} from [37, AppendixB], WEST^{nso} has the distinction between the absolute equality $=$ and the set-theoretic (but still intensional) equality $\approx$, and the replacement operator $\mathbb{R}$ applies only to those operations on sets which respect the latter equality $\approx$. This restriction of $\mathbb{R}$ requires a more complex realizability notion than in WEST^{int}, since, in order to realize collection scheme, we need to require a realizer of $(\forall u \in x)(\exists v)A[u, v]$ always respect $\approx$, as discussed in Subsection 1.6.

Definition 30 (Realizing relation $\mathfrak{er}$)**.** For each $\mathcal{L}_\in$ formula A in which neither f nor g occurs, the $\mathcal{L}_\in^{nso}$ formulae $f \,\mathfrak{er}\, A$, read as "f *realizes* A" or "f is a *realizer* of A", and $f \sim_A g$, read as "f and g are *equivalent realizers* of A", are defined by simultaneous induction on A as follows:

$$
\begin{aligned}
f \,\mathfrak{er}\, (x = y) &:\equiv (f = f) \wedge (x \approx y); & f \sim_{x=y} g &:\equiv f = g;\\
f \,\mathfrak{er}\, (x \in y) &:\equiv (f = f) \wedge (x \in y); & f \sim_{x\in y} g &:\equiv f = g;\\
f \,\mathfrak{er}\, \bot &:\equiv \bot & f \sim_{\bot} g &:\equiv f = f \wedge g = g;\\
f \,\mathfrak{er}\, (B \wedge C) &:\equiv (\mathbf{p}_0(f) \,\mathfrak{er}\, B) \wedge (\mathbf{p}_1(f) \,\mathfrak{er}\, C); & f \sim_{B\wedge C} g &:\equiv (\mathbf{p}_0(f) \sim_B \mathbf{p}_0(g)) \wedge (\mathbf{p}_1(f) \sim_C \mathbf{p}_1(g));
\end{aligned}
$$

$$
\begin{aligned}
f \,\mathfrak{er}\, (B \to C) &:\equiv f = f \wedge (\forall u, v \,\mathfrak{er}\, B)(f(u){\downarrow} \wedge f(u) \,\mathfrak{er}\, C \wedge (u \sim_B v \to f(u) \sim_C f(v)));\\
&\qquad f \sim_{B\to C} g :\equiv (f = f \wedge g = g) \wedge (\forall u \,\mathfrak{er}\, B)(f(u) \sim_C g(u)));\\
f \,\mathfrak{er}\, (\forall x B[x]) &:\equiv (\forall u, v \in \mathcal{S})(f(u){\downarrow} \wedge f(u) \,\mathfrak{er}\, B[u] \wedge (u \approx v \to f(u) \sim_{B[u]} f(v)));\\
&\qquad f \sim_{\forall x B[x]} g :\equiv (f = f \wedge g = g) \wedge (\forall u \in \mathcal{S})(f(u) \sim_{B[u]} g(u));\\
f \,\mathfrak{er}\, (B \vee C) &:\equiv (\mathbf{p}_0(f) = 0 \wedge \mathbf{p}_1(f) \,\mathfrak{er}\, B) \vee (\mathbf{p}_0(f) = 1 \wedge \mathbf{p}_1(f) \,\mathfrak{er}\, C);\\
&\qquad f \sim_{B\vee C} g :\equiv (\mathbf{p}_0(f) = \mathbf{p}_0(g) = 0 \wedge \mathbf{p}_1(f) \sim_B \mathbf{p}_1(g))\\
&\qquad\qquad \vee (\mathbf{p}_0(f) = \mathbf{p}_0(g) = 1 \wedge \mathbf{p}_1(f) \sim_C \mathbf{p}_1(g));\\
f \,\mathfrak{er}\, (\exists x B[x]) &:\equiv \mathcal{S}(\mathbf{p}_0(f)) \wedge (\mathbf{p}_1(f) \,\mathfrak{er}\, B[\mathbf{p}_0(f)]); \quad f \sim_{\exists x B[x]} g :\equiv (\mathbf{p}_0(f) \approx \mathbf{p}_0(g)) \wedge (\mathbf{p}_1(f) \sim_{B[\mathbf{p}_0(f)]} \mathbf{p}_1(g)).
\end{aligned}
$$

We say that an $\mathcal{L}_\in$ formula $A[\vec{x}]$ with at most the free variables $\vec{x}$ is *realizable in* a theory, if the formula $\exists f(f \,\mathfrak{er}\, \forall x A[\vec{x}])$ is provable in the theory. We just say that a formula is *realizable* if it is realizable in WEST_0^{nso}. For $\mathcal{L}_\in^{nso}$ terms $\vec{t}$, we write $f \,\mathfrak{er}\, A[\vec{t}]$ for $(f \,\mathfrak{er}\, A[\vec{u}])[\vec{t}/\vec{u}]$.

It is easy to see by induction on a formula $A[\vec{x}]$ that $\vec{u} \approx \vec{v}$ implies the equivalences between $f \,\mathfrak{er}\, A[\vec{u}]$ and $f \,\mathfrak{er}\, A[\vec{v}]$, and between $f \sim_{A[\vec{u}]} g$ and $f \sim_{A[\vec{v}]} g$, as stated in the next lemma. Therefore, in the definition of $f \,\mathfrak{er}\, (\forall x B[x])$, $f(u) \sim_{B[u]} f(v)$ is equivalent to $f(u) \sim_{B[v]} f(v)$, and in the definition of $f \sim_{\exists x B[x]} g$, $\mathbf{p}_1(f) \sim_{B[\mathbf{p}_0(f)]} \mathbf{p}_1(g)$ is equivalent to $\mathbf{p}_1(f) \sim_{B[\mathbf{p}_0(g)]} \mathbf{p}_1(g)$.

Also, the free variables of $f \,\mathfrak{er}\, A$ are the variable f and the free variables of A.

This realizability is called *extensional realizability*, because it is based on the equation in the sense of functional extensionality. Notice however that the set-theoretic axiom of extensionality is *not* realized.

Lemma 31. WEST_0^{nso} proves (i) $(t \,\mathfrak{er}\, A) \to t{\downarrow}$; (ii) $(s \sim_A t) \to t{\downarrow} \wedge s{\downarrow}$; (iii) $f \sim_A f$; (iv) $s \sim_A t \to t \sim_A s$; (v) $t_1 \sim_A t_2 \wedge t_2 \sim_A t_3 \to t_1 \sim_A t_3$; (vi) $t \approx s \wedge f \,\mathfrak{er}\, A[t] \to f \,\mathfrak{er}\, A[s]$; and (vii) $t \sim_A s \wedge (t \,\mathfrak{er}\, A) \to (s \,\mathfrak{er}\, A)$.

Proof. By induction on A. Especially (i) and (ii) are from the seemingly redundant clauses $f = f$ and $g = g$. □

From the viewpoint that the realizability machinery is a "miniature" of Martin-Löf type theory, the use of $\sim_A$ can be seen as a use of identity type. In the usual realizability interpretation, the "miniatures" of identity types can be taken as the trivial ones. For our purpose, however, we have to take non-trivial ones, and, accordingly, we have to define the type structure (in the sense of Martin-Löf type theory, not that of explicit mathematics) in a more complex way, e.g., the realizers of $A \to B$ (corresponding to the inhabitants of type $A \to B$) are those operators which assign a realizer of B to any realizer of A respecting the "miniatures" of identity types. However, identity types are trivialized in another sense: according to the analogy, any object is an inhabitant of the identity type $f \sim_A g$ whenever $f \sim_A g$ holds, and no inhabitants exist otherwise, and the inhabitant of the type $(f \sim_A g) \to (hf \sim_B hg)$ induced by $h : A \to B$ is identity.

Actually we could enhance this realizability to an interpretation of a kind of Martin-Löf type theory with identity type, by enhancing the realizability of $\to$-type formulae to the Π-type and by choosing appropriate base types. Based on a similar idea, Griffor and Rathjen [15] defined an interpretation of a family of Martin-Löf type theories in variants of KP. However, the interpretation of a type (in the sense of type theory) or the class of realizers of a fixed formula does not necessarily form a set, the heir of a type in the sense of explicit mathematics. Because of this difference, the base type associated with $\mathcal{S}$ cannot be the V-type.

By requiring the witness of membership $\in$ (namely, r realizes $a \in b$ if $\langle r, a\rangle \in b$), the realizability of Tupailo [45] allows the class of realizers of a Δ_0 formulae to form a type in explicit mathematics, and hence it realizes the separation scheme for non-negative Δ_0 formulae, while ours realizes only that for negative Δ_0. On the other hand, Tupailo's realizability prevents us from having canonical realizers, which we will need later (especially in Lemma 40 where the realizability of N-Beta, rather than that of Beta itself, is important; see [37, Subsection A.4]).

5.3. *Canonical realizer for negative formulae*

Definition 32 (Canonical realizer)**.** We assign to each finite sequence $\vec{x} = x_0, ..., x_n$ of variables and negative $\mathcal{L}$ formula $A[\vec{x}]$, in which at most the variables $\vec{x}$ occur freely, a closed $\mathcal{L}_\in^{nso}$ term $c_{A,\vec{x}}$ inductively:

$$c_{A,\vec{x}} := \lambda\vec{x}.\emptyset \text{ if } A \text{ is atomic}; \qquad c_{B\wedge C,\vec{x}} := \langle c_{B,\vec{x}}, c_{C,\vec{x}}\rangle;$$
$$c_{B\to C,\vec{x}} := \lambda\vec{x}, g.c_{C,\vec{x}}(\vec{x}); \text{ where } g \text{ is fresh} \qquad c_{\forall yB,\vec{x}} := \lambda\vec{x}, y.c_{B,\vec{x},y}(\vec{x}, y),$$

It is easy to see that $c_{A,\vec{x}}$ defined above is a closed term, and $c_{A,\vec{x}}(\vec{s}) = c_{A,\vec{x}}(\vec{t})$.

Lemma 33. For a negative $\mathcal{L}_\in$ formula A, WEST_0^{nso} proves the following, where $A^\approx$ is as in Definition 15:

(i) $A^\approx[\vec{y}] \leftrightarrow c_{A,\vec{x}}(\vec{y})\,\mathfrak{er}\,A[\vec{y}]$, and
(ii) $f\,\mathfrak{er}\,A[\vec{x}] \to A^\approx[\vec{x}]$,

Proof. We prove the two statements by induction on the negative formula A.

1. If A is atomic, the statements are obvious.
2. Let us consider the case of $A \equiv B \wedge C$.
   (i) By definition $c_{B\wedge C,\vec{x}}(\vec{y})\,\mathfrak{er}\,(B[\vec{y}] \wedge C[\vec{y}])$ is equivalent to the conjunction of $c_{B,\vec{x}}(\vec{y})\,\mathfrak{er}\,B[\vec{y}]$ and $c_{C,\vec{x}}(\vec{y})\,\mathfrak{er}\,C[\vec{y}]$, and hence, by induction hypothesis, to $B^\approx[\vec{y}] \wedge C^\approx[\vec{y}]$.
   (ii) $f\,\mathfrak{er}\,B[\vec{y}] \wedge C[\vec{y}]$ is, by definition, the conjunction of $\mathbf{p}_0(f)\,\mathfrak{er}\,B[\vec{y}]$ and $\mathbf{p}_1(f)\,\mathfrak{er}\,C[\vec{y}]$, which with induction hypothesis imply $B^\approx[\vec{y}]$ and $C^\approx[\vec{y}]$, that is $A^\approx[\vec{y}]$.
3. Let $A \equiv B \to C$.
   (i) By definition, $c_{B\to C,\vec{x}}(\vec{y})\,\mathfrak{er}\,(B[\vec{y}] \to C[\vec{y}])$ means that, for any g, $g\,\mathfrak{er}\,B[\vec{y}]$ implies $c_{C,\vec{x}}(\vec{y})\,\mathfrak{er}\,C[\vec{y}]$, since $c_{B\to C,\vec{x}}(\vec{y})$ is a constant operation and therefore preserves the equality $\sim$ trivially.
   If $B^\approx[\vec{y}] \to C^\approx[\vec{y}]$ holds, then for any g, $g\,\mathfrak{er}\,B[\vec{y}]$, with induction hypothesis, implies $B^\approx[\vec{y}]$ and by assumption $C^\approx[\vec{y}]$, which implies $c_{B\to C,\vec{x}}(\vec{y})(g) = c_{C,\vec{x}}(\vec{y})\,\mathfrak{er}\,C[\vec{y}]$ again by induction hypothesis.
   The converse is a special case of (ii) proved below.
   (ii) Assume $f\,\mathfrak{er}\,(B[\vec{y}] \to C[\vec{y}])$, particularly, for any g, $g\,\mathfrak{er}\,B[\vec{y}]$ implies $f(g)\,\mathfrak{er}\,C[\vec{y}]$. To show that $B^\approx[\vec{y}] \to C^\approx[\vec{y}]$ holds, assume $B^\approx[\vec{y}]$. Then by induction hypothesis $c_{B,\vec{x}}(\vec{y})\,\mathfrak{er}\,B[\vec{y}]$ and so $f(c_{B,\vec{x}}(\vec{y}))\,\mathfrak{er}\,C[\vec{y}]$, which with induction hypothesis implies $C^\approx[\vec{y}]$.
4. Let $A[\vec{x}] \equiv \forall zB[\vec{y}, z]$.
   (i) By definition, $c_{\forall zB,\vec{x}}(\vec{y})\,\mathfrak{er}\,\forall zB[\vec{y}, z]$ means that $c_{B,\vec{x},z}(\vec{y}, z)\,\mathfrak{er}\,B[\vec{y}, z]$ for any $z \in \mathcal{S}$, since $c_{B,\vec{x},z}$ is a constant operation. Thus, by induction hypothesis, $c_{\forall zB,\vec{x}}(\vec{y})\,\mathfrak{er}\,\forall zB[\vec{y}, z]$ is equivalent to $(\forall z \in \mathcal{S})B^\approx[\vec{y}, z]$, i.e., $A^\approx[\vec{y}]$.
   (ii) Assume $f\,\mathfrak{er}\,\forall zB[\vec{y}, z]$. Then, for any $z \in \mathcal{S}$, $f(z)\,\mathfrak{er}\,B[\vec{y}, z]$ holds and hence by induction hypothesis $(\forall z \in \mathcal{S})(B^\approx[\vec{y}, z])$. ☐

Corollary 34. For any negative $\mathcal{L}_\in$ formula $A[\vec{x}]$, WEST_0^{nso} proves $A^\approx[\vec{x}] \leftrightarrow \exists f(f\,\mathfrak{er}\,A[\vec{x}])$.

5.4. Realizing IKP$^-$

Lemma 35. If $A[\vec{x}]$ follows intuitionistically from $B_0[\vec{x}]$, ..., $B_n[\vec{x}]$, then WEST_0^{nso} proves $\exists f(f \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to A[\vec{x}]))$. Thus, realizability in WEST_0^{nso} is closed under inferences of intuitionistic logic.

Proof. By induction on a proof of $A[\vec{x}]$ from $B_0[\vec{x}]$, ..., $B_n[\vec{x}]$ in Hilbert system of intuitionistic logic.

It is almost trivial that all the logical axioms of intuitionistic logic are realizable, including equality axioms which are all negative. We check this only in the cases of $C[\vec{x}, z] \to \exists z C[\vec{x}, z]$ and $\forall z C[\vec{x}, z] \to C[\vec{x}, z]$.

$$t_1 := \lambda\vec{x}, z, f.\langle z, f\rangle \text{ and } t_2 := \lambda\vec{x}, z, g.g(z)$$

satisfy $t_1 \,\mathfrak{er}\, \forall\vec{x}, z(C[\vec{x}, z] \to \exists z C[\vec{x}, z])$ and $t_2 \,\mathfrak{er}\, \forall\vec{x}, z(\forall z C[\vec{x}, z] \to C[\vec{x}, z])$, since $\vec{x} \approx \vec{y}$, $z_1 \approx z_2$, $f_1 \sim_{C[\vec{x}, z_1]} f_2$ and $g_1 \sim_{\forall z C[\vec{x}, z]} g_2$ imply $\langle z_1, f_1\rangle \sim_{\exists z C[\vec{x}, z]} \langle z_2, f_2\rangle$ and $g_1(z_1) \sim_{C[\vec{x}, z_1]} g_2(z_2)$.

Let us turn to the rules. Assume first that $A[\vec{x}]$ is the conclusion of Modus Ponens whose premises are $C[\vec{x}]$ and $C[\vec{x}] \to A[\vec{x}]$. The induction hypotheses give us f and g such that

$$f \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to C[\vec{x}]) \text{ and } g \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to C[\vec{x}] \to A[\vec{x}]).$$

Then $t := \lambda\vec{x}, h_0, ..., h_n.g(\vec{x}, h_0, ..., h_n, f(\vec{x}, h_0, ..., h_n))$ satisfies $t \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to A[\vec{x}])$, since $\vec{x} \approx \vec{y}$ and $h_0 \sim_{B_0[\vec{x}]} j_0$, ..., $h_n \sim_{B_n[\vec{x}]} j_n$ imply both $f(\vec{x}, h_0, ..., h_n) \sim_{C[\vec{x}]} f(\vec{y}, j_0, ..., j_n)$ and $g(\vec{x}, h_0, ..., h_n) \sim_{C[\vec{x}] \to A[\vec{x}]} g(\vec{y}, j_0, ..., j_n)$, from which follows $g(\vec{x}, h_0, ..., h_n, f(\vec{x}, h_0, ..., h_n)) \sim_{A[\vec{x}]} g(\vec{y}, j_0, ..., j_n, f(\vec{y}, j_0, ..., j_n))$.

Next assume that $A[\vec{x}] \equiv C_1[\vec{x}] \to \forall z C_2[\vec{x}, z]$ is the conclusion of $\forall$-rule whose only premise is $C_1[\vec{x}] \to C_2[\vec{x}, z]$. By the eigenvariable condition z does not occur in $C_1[\vec{x}]$, and we may assume that z does not occur in $B_0[\vec{x}]$, ..., $B_n[\vec{x}]$ either. The induction hypothesis gives us f such that

$$f \,\mathfrak{er}\, \forall\vec{x}, z(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to C_1[\vec{x}] \to C_2[\vec{x}, z]).$$

Then $t := \lambda\vec{x}, h_0, ..., h_n, h_{n+1}, z.f(\vec{x}, z, h_0, ..., h_n, h_{n+1})$ satisfies $t \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to C_1[\vec{x}] \to \forall z C_2[\vec{x}, z]))$, since if all of $\vec{x} \approx \vec{y}$, $h_0 \sim_{B_0[\vec{x}]} j_0$, ..., $h_n \sim_{B_n[\vec{x}]} j_n$, $h_{n+1} \sim_{C_1[\vec{x}]} j_{n+1}$ and $z_1 \approx z_2$ hold then $f(\vec{x}, z_1, h_0, ..., h_n, h_{n+1}) \sim_{C_2[\vec{x}, z_1]} f(\vec{y}, z_2, j_0, ..., j_n, j_{n+1})$.

Finally assume that $A[\vec{x}] \equiv \exists z C_1[\vec{x}, z] \to C_2[\vec{x}]$ is the conclusion of $\exists$-rule whose only premise is $C_1[\vec{x}, z] \to C_2[\vec{x}]$. By the eigenvariable condition z does not occur in $C_2[\vec{x}]$, and we may assume that z does not occur in $B_0[\vec{x}]$, ..., $B_n[\vec{x}]$ either. The induction hypothesis gives us f such that

$$f \,\mathfrak{er}\, \forall\vec{x}, z(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to C_1[\vec{x}, z] \to C_2[\vec{x}]).$$

Then $t := \lambda\vec{x}, h_0, ..., h_n, h_{n+1}.f(\vec{x}, \mathbf{p}_0(h_{n+1}), h_0, ..., h_n, \mathbf{p}_1(h_{n+1}))$ satisfies $t \,\mathfrak{er}\, \forall\vec{x}(B_0[\vec{x}] \to ... \to B_n[\vec{x}] \to \exists z C_1[\vec{x}, z] \to C_2[\vec{x}]))$, since if all of $\vec{x} \approx \vec{y}$, $h_0 \sim_{B_0[\vec{x}]} j_0$, ..., $h_n \sim_{B_n[\vec{x}]} j_n$, and $h_{n+1} \sim_{\exists z C_1[\vec{x}, z]} j_{n+1}$ hold then $\mathbf{p}_0(h_{n+1}) \approx \mathbf{p}_0(j_{n+1})$ and $\mathbf{p}_1(h_{n+1}) \sim_{C_1[\vec{x}, \mathbf{p}_0(h_{n+1})]} \mathbf{p}_1(j_{n+1})$, and so $f(\vec{x}, \mathbf{p}_0(h_{n+1}), h_0, ..., h_n, \mathbf{p}_1(h_{n+1})) \sim_{C_2[\vec{x}]} f(\vec{x}, \mathbf{p}_0(j_{n+1}), j_0, ..., j_n, \mathbf{p}_1(j_{n+1}))$. $\square$

Theorem 36. All the axioms (and so theorems) of IKP_0^- and full collection scheme are realizable in WEST_0^{nso}.

Proof. All the axioms of IKP_0^-, except Δ_0 collection, are of the from $\exists y A[\vec{x}, y]$ (or $A[\omega]$, which is included in the former case) with A being negative. Since there is a closed $\mathcal{L}_\in^{nso}$ term t such that WEST_0^{nso} proves

$$(\forall\vec{x}, \vec{y} \in \mathcal{S})(t(\vec{x}) \in \mathcal{S} \wedge A^{\approx}[\vec{x}, t(\vec{x})] \wedge (\vec{x} \approx \vec{y} \to t(\vec{x}) \approx t(\vec{y}))),$$

where $A^{\approx}$ is as in Lemma 33, thus $\lambda\vec{x}.\langle t(\vec{x}), c_{A,\vec{x},y}(\vec{x}, t(\vec{x}))\rangle$ realizes $\exists y A[\vec{x}, y]$.

For collection scheme, assume $g \,\mathfrak{er}\, (\forall u \in x)\exists v A[u, v, \vec{z}]$. Then, since $\emptyset \,\mathfrak{er}\, u \in x$ for $u \in x$,

$$\begin{aligned}&(\forall u \in x)(g(u, \emptyset)\downarrow \wedge\, \mathbf{p}_0(g(u, \emptyset)) \in \mathcal{S} \wedge \mathbf{p}_1(g(u, \emptyset)) \,\mathfrak{er}\, A[u, \mathbf{p}_0(g(u, \emptyset)), \vec{z}])\\ &\wedge (\forall u_1, u_2 \in x)(u_1 \approx u_2 \to \mathbf{p}_0(g(u_1, \emptyset)) \approx \mathbf{p}_0(g(u_2, \emptyset)) \wedge \mathbf{p}_1(g(u_1, \emptyset)) \sim_{A[u_1, \mathbf{p}_0(g(u_1, \emptyset)), \vec{z}]} \mathbf{p}_1(g(u_2, \emptyset))).\end{aligned}$$

Let $f := \lambda g, u.\mathbf{p}_0(g(u, \emptyset))$. Then $f(g)$ satisfies the hypothesis of the axiom for $\mathbb{R}$, i.e.,

$$(\forall u \in x)(f(g, u)\downarrow \wedge f(g, u) \in \mathcal{S}) \text{ and } (\forall u_1, u_2 \in x)(u_1 \approx u_2 \to f(g, u_1) \approx f(g, u_2)).$$

Then we have $(\forall u \in x)(f(g,u) \in \mathbb{R}(x, f(g)))$ and $t[g,u] \,\mathfrak{er}\, (\exists v \in \mathbb{R}(x, f(g)))A[u,v,\vec{z}])$ for any $x \in u$, where

$$t[g,u] := \langle f(g,u), \langle \emptyset, \mathbf{p}_1(g(u,\emptyset))\rangle\rangle = \langle \mathbf{p}_0(g(u,\emptyset)), \langle \emptyset, \mathbf{p}_1(g(u,\emptyset))\rangle\rangle.$$

Now $\lambda h.t[g,u]$ realizes $(x \in u \to (\exists v \in \mathbb{R}(x, f(g)))A[u,v,\vec{z}])$ because it is a constant operation. Moreover, $u_1 \approx u_2$ implies $f(g,u_1) \approx f(g,u_2)$ as shown before and so $\lambda h.t[g,u_1] \sim_{(u_1 \in x \to (\exists v \in \mathbb{R}(a,f(g)))A[u_1,v,\vec{z}])} \lambda h.t[g,u_2]$. Thus $\lambda u,h.t[g,u] \,\mathfrak{er}\, (\forall u \in x)(\exists v \in \mathbb{R}(x, f(g)))A[u,v,\vec{z}]$ and so

$$\langle \mathbb{R}(x, f(g)), \lambda u,h.t[g,u]\rangle \,\mathfrak{er}\, (\exists y)(\forall u \in x)(\exists v \in y)A[u,v,\vec{z}].$$

Now, it remains to show that the conjunction of $x_1 \approx x_2$, $\vec{z}_1 \approx \vec{z}_2$ and $g_1 \sim_{(\forall u \in x_1)\exists v A[u,v,\vec{z}_1]} g_2$ implies $\langle \mathbb{R}(x_1, f(g_1)), \lambda u,h.t[g_1,u]\rangle \sim_{(\exists y)(\forall u \in x_1)(\exists v \in y)A[u,v,\vec{z}_1]} \langle \mathbb{R}(x_2, f(g_2)), \lambda u,h.t[g_2,u]\rangle$, i.e.,

$$\mathbb{R}(x_1, f(g_1)) \approx \mathbb{R}(x_2, f(g_2)) \text{ and } t[g_1,u] \sim_{(\exists v \in \mathbb{R}(x_1, f(g)))A[u,v,\vec{z}_1]} t[g_2,u] \text{ for all } u \in x_1.$$

For the former, by axiom (B.6), it suffices to show $f(g_1,u) \approx f(g_2,u)$, i.e., $\mathbf{p}_0(g_1(u,\emptyset)) \approx \mathbf{p}_0(g_2(u,\emptyset))$ for any $u \in x_1$, which follows from $g_1 \sim_{(\forall u \in x_1)\exists v A[u,v,\vec{z}_1]} g_2$. For the latter, we need to show, for any $u \in x_1$, $f(g_1,u) \approx f(g_2,u)$, which we have just shown; and $\mathbf{p}_1(g_1(u,\emptyset)) \sim_{A[u,f(g_1,u)]} \mathbf{p}_1(g_2(u,\emptyset))$ which follows from $g_1 \sim_{(\forall u \in x_1)\exists v A[u,v,\vec{z}_1]} g_2$, the assumption.

What we have shown is that $\lambda \vec{z},x,g.\langle \mathbb{R}(x, f(g)), \lambda u,h.t[g,x,u]\rangle$ realizes the instance of collection scheme $(\forall \vec{z},x)((\forall u \in x)\exists v A[u,v,\vec{z}] \to (\exists y)(\forall u \in x)(\exists v \in y)A[u,v,\vec{z}])$. ☐

Lemma 37. All the axioms (and so theorems) of IKP_r^- are realizable in WEST_r^{nso}.

Proof. For negative formula $A[x]$, the instance $(\forall x)((\forall y \in x)A[x] \to A[x]) \to \forall x A[x]$ of foundation scheme is itself negative and therefore it is realizable whenever it holds, by Corollary 34. ☐

Lemma 38. All the instances of foundation scheme are realizable in WEST^{nso}; and all the instances of foundation scheme on ω are realizable in $\mathsf{WEST}_\omega^{nso}$.

Proof. First we are working in WEST^{nso}. Let us use the following abbreviation:

$$\mathrm{ExtOp}^A[f,x] :\equiv (\forall z)(x \approx z \wedge (f(x)\downarrow \vee f(z)\downarrow) \to f(x) \sim_{A[x]} f(z)).$$

Assume $g \,\mathfrak{er}\, (\forall x)((\forall y \in x)A[y] \to A[x])$. By the definition of $\mathfrak{er}$, $x_1 \approx x_2 \to g(x_1) \sim_{(\forall y \in x_1)A[y] \to A[x_1]} g(x_2)$ holds for any x_1 and x_2, and $(\forall h)((h \,\mathfrak{er}\, (\forall y \in x)A[y]) \to g(x,h) \,\mathfrak{er}\, A[x])$ for any x.

Assume also $(\forall y \in x)\mathrm{ExtOp}^A[f,y]$ and $(\forall y \in x)(f(y) \,\mathfrak{er}\, A[y])$. Then $(\forall y)(\lambda u.f(y) \,\mathfrak{er}\, (y \in x \to A[y]))$ and, by $(\forall y \in x)\mathrm{ExtOp}^A[f,y]$, also $\lambda y,u.f(y) \,\mathfrak{er}\, (\forall y \in x)A[y]$ hold. Thus we have $g(x, \lambda y,u.f(y)) \,\mathfrak{er}\, A[x]$. Moreover, if $x \approx z$, then $g(x) \sim_{(\forall y \in x)A[y] \to A[x]} g(z)$ and so $g(x, \lambda y,u.f(y)) \sim_{A[x]} g(z, \lambda y,u.f(y))$. What we have seen is that, for any f,

$$(\forall y \in x)(\mathrm{ExtOp}^A[f,y] \wedge f(y) \,\mathfrak{er}\, A[y]) \to \mathrm{ExtOp}^A[\lambda z.g(z, \lambda y,u.f(y)), x] \wedge g(x, \lambda y,u.f(y)) \,\mathfrak{er}\, A[x].$$

Therefore, by setting $t[g] = \mathbf{fix}(\lambda f,x.g(x, \lambda y,u.f(y)))$ we have $(\forall x)(t[g](x) \simeq g(x, \lambda y,u.t[g](y)))$, and hence

$$(\forall y \in x)(\mathrm{ExtOp}^A[t[g],y] \wedge t[g](y) \,\mathfrak{er}\, A[y]) \to \mathrm{ExtOp}^A[t[g],x] \wedge t[g](x) \,\mathfrak{er}\, A[x].$$

By foundation, $(\forall x)(\mathrm{ExtOp}^A[t[g],x] \wedge t[g](x) \,\mathfrak{er}\, A[x])$, which implies $(\forall x)(t[g](x)\downarrow)$. The argument at the beginning of this paragraph shows $\lambda y,u.t[g](y) \,\mathfrak{er}\, (\forall y \in x)A[y]$.

To conclude $t[g] \,\mathfrak{er}\, (\forall x)A[x]$, we need to show that $x_1 \approx x_2$ implies $t[g](x_1) \sim_{A[x_1]} t[g](x_2)$. The latter is equivalent to $g(x_1, \lambda y,u.t[g](y)) \sim_{A[x_1]} g(x_2, \lambda y,u.t[g](y))$, which is implied by $x_1 \approx x_2$, by the assumption $g \,\mathfrak{er}\, (\forall x)((\forall y \in x)A[y] \to A[x])$ and by $\lambda y,u.t[g](y) \,\mathfrak{er}\, (\forall y \in x)A[y]$.

Finally, to conclude $\lambda g.t[g] \,\mathfrak{er}\, ((\forall x)((\forall y \in x)A[y] \to A[x]) \to \forall x A[x])$, it remains to show that $t[g_1] \sim_{\forall x A[x]} t[g_2]$ is implied by $g_1 \sim_{(\forall x)((\forall y \in x)A[y] \to A[x])} g_2$. Assume the latter. It suffices to prove $t[g_1](x) \sim_{A[x]} t[g_2](x)$ by foundation on x. The induction hypothesis means $(\forall y \in x)(t[g_1](y) \sim_{A[y]} t[g_2](y))$ which implies

$(\forall y)(\lambda u.t[g_1](y) \sim_{y\in x\to A[y]} \lambda u.t[g_2](y))$ since $\sim_{y\in x\to A[y]}$ is trivial if $y \notin x$, and hence $\lambda y, u.t[g_1](y) \sim_{(\forall y\in x)A[y]} \lambda y, u.t[g_2](y)$. The assumption implies $g_1(x) \sim_{(\forall y\in x)A[y]\to A[x]} g_2(x)$ for any x and hence

$$t[g_1](x) = g_1(x, \lambda y, u.t[g_1](y)) \sim_{A[x]} g_2(x, \lambda y, u.t[g_2](y)) = t[g_2](x).$$

Similarly we can show that the foundation on ω can be realized in $\mathsf{WEST}^{nso}_{\omega}$. □

Corollary 39. All the axioms (and so theorems) of IKP^-, of IKP^-_w and of IKP^-_ω are realizable in WEST^{nso}, in WEST^{nso}_w and in $\mathsf{WEST}^{nso}_\omega$, respectively.

The following lemma is for the extension by Axiom Beta. Notice that $\mathrm{Clp}[f, b, r]$ (as well as $\mathrm{DwCl}[b, a, r]$ and $\mathrm{Prog}[b, a, r]$) is an $\mathcal{L}_\in$ formula, since it does not contain any occurrences of $\approx$ (the definition contains $\approx$ only for abbreviation).

Lemma 40. The following is realizable in WEST^{nso}_0 plus Axiom Beta (C.14):

(IKP^-.9) N-Beta $(\forall a, r)(\exists f, b)(\mathrm{DwCl}[b, a, r] \wedge \mathrm{Prog}[b, a, r] \wedge \mathrm{Fun}[f] \wedge \mathrm{Dom}[f, b] \wedge \mathrm{Clp}[f, b, r])^N$.

Proof. By the same argument as the first part of the proof of Theorem 36 and by the fact that WEST^{nso}_0 is on classical logic, we can prove this. □

6. Forcing and Negative Interpretations

The interpretation of intensional variants of KP in those of IKP^- was already given in Sato and Zumbrunnen [37, Sections 5 and 6], which is a modification of Avigad's [3] interpretation of KP in IKP, the intuitionistic version of KP. Here we only briefly summarize the results proved there.

6.1. Semi-constructive set theory $\mathsf{IKP}^\sharp + (\Delta^{s-}_0\text{-}\mathsf{MP})$

We need one more family of systems, generally weaker than the corresponding variants of IKP^-.

Definition 41 ($\mathsf{IKP}^\sharp$ and (Δ^{s-}_0-MP))**.** The $\mathcal{L}_\in$-theory $\mathsf{IKP}^\sharp$ is based on intuitionistic first-order logic with equality axioms and consists of the following non-logical axioms.

($\mathsf{IKP}^\sharp$.1) $\exists x((\forall y \in a)\neg\neg(y \in x) \wedge (\forall y \in x)(\forall z \in y)\neg\neg(z \in x))$ (N-transitive superset).

($\mathsf{IKP}^\sharp$.2) $\exists x(\neg\neg a \in x \wedge \neg\neg b \in x)$ (N-pairing).

($\mathsf{IKP}^\sharp$.3) $\exists x(\forall y \in a)(\forall z \in y)\neg\neg(z \in x)$ (N-union).

($\mathsf{IKP}^\sharp$.4) $\exists x((\forall y \in x)(\neg\neg y \in a \wedge A[y]) \wedge (\forall y \in a)(A[y] \to \neg\neg y \in x))$ for all strongly negative Δ_0 formulae $A[y]$ in which x does not occur ($\Delta^{\mathrm{s}-}_0$ N-separation).

($\mathsf{IKP}^\sharp$.5) $(\forall x \in a)\exists y A[x, y] \to \exists z(\forall x \in a)\neg(\forall y \in z)\neg A[x, y]$ for all strongly negative Δ_0 formulae $A[x, y]$ in which z does not occur ($\Delta^{\mathrm{s}-}_0$ collection$^\sharp$).

($\mathsf{IKP}^\sharp$.6) $\forall x((\forall y \in x)A[y] \to A[x]) \to \forall x A[x]$ for all strongly negative formulae $A[x]$ ($\mathcal{L}^{\mathrm{s}-}_\in$-$\mathsf{Ind}$).

($\mathsf{IKP}^\sharp$.7) $\mathrm{Ind}[\omega]^N \wedge (\forall x \subseteq \omega)(\mathrm{Ind}[x] \to \omega \subseteq x)^N \wedge (\forall y)(\exists z)\mathrm{famfun}[\omega, y, z]^N$ (N-infinity), where $\mathrm{Ind}[x]$ and $\mathrm{famfun}[x, y, z]$ are defined as in the definition of IKP^- in Definition 29.

$\mathsf{IKP}^\sharp_0$ is $\mathsf{IKP}^\sharp$ minus ($\mathcal{L}^{\mathrm{s}-}_\in$-$\mathsf{Ind}$); $\mathsf{IKP}^\sharp_\omega$ is $\mathsf{IKP}^\sharp_0$ plus $\in$-induction on ω for all *strongly* negative $\mathcal{L}_\in$ formulae; $\mathsf{IKP}^\sharp_r$ is $\mathsf{IKP}^\sharp_0$ plus $\in$-induction for all *strongly* negative Δ_0 formulae; and $\mathsf{IKP}^\sharp_w$ is $\mathsf{IKP}^\sharp_\omega$ plus $\mathsf{IKP}^\sharp_r$.

These theories will be augmented by:

(Δ^{s-}_0-MP) $\neg\forall x A[x] \to \exists y\neg(\forall x \in y)A[x]$, for all strongly negative Δ_0 formulae $A[x]$ of $\mathcal{L}_\in$.

We use variants of $\mathsf{IKP}^\sharp$ only with (Δ^{s-}_0-MP), a set-theoretic version of Markov's principle which is often called semi-constructive. The resulting theories (e.g., $\mathsf{IKP}^\sharp + (\Delta^{s-}_0\text{-}\mathsf{MP})$) are also called semi-constructive.

6.2. Avigad forcing

Definition 42 ($\mathrm{Tr}_{\mathfrak{S}}$)**.** Let $\mathfrak{S}$ be a finite sequence $D_0[z,\vec{y}],...,D_{n-1}[z,\vec{y}]$ of strongly negative Δ_0 formulae with at most the variables z, $\vec{y}=y_0,...,y_m$ free. $\mathrm{Tr}_{\mathfrak{S}}[p,u]$ is a strongly negative Δ_0 formula equivalent to

$$\bigwedge\nolimits_{i=0}^{n-1} \forall\vec{y}(\langle i,y_0,...,y_m\rangle \in p \to (\forall z\in u)D_i[z,\vec{y}]) \quad \wedge \quad \bigwedge\nolimits_{i=n}^{2n-1} \forall z\forall\vec{y}(\langle i,z,y_0,...,y_m\rangle \in p \to D_{i-n}[z,\vec{y}]).$$

Definition 43 ($\Vdash_{\mathfrak{S}}$)**.** Let $\mathfrak{S}$ be a finite sequence of strongly negative Δ_0 formulae with at most the variables z, $\vec{y}$ free. For an arbitrary $\mathcal{L}_\in$ formula A, the $\mathcal{L}_\in$ formula $p\Vdash_{\mathfrak{S}} A$ is defined inductively as follows:

$p\Vdash_{\mathfrak{S}} A :\equiv \exists u(\mathrm{Tr}_{\mathfrak{S}}[p,u]\to\neg\neg A)$ if A is atomic; $p\Vdash_{\mathfrak{S}} B\wedge C :\equiv (p\Vdash B)\wedge(p\Vdash C)$;
$p\Vdash_{\mathfrak{S}} B\vee C :\equiv (p\Vdash B)\vee(p\Vdash C)$; $p\Vdash_{\mathfrak{S}} B\to C :\equiv (\forall q\supseteq p)((q\Vdash B)\to(q\Vdash C))$;
$p\Vdash_{\mathfrak{S}} \forall xB[x] :\equiv \forall v(p\Vdash_{\mathfrak{S}} B[v])$; $p\Vdash_{\mathfrak{S}} \exists xB[x] :\equiv \exists v(p\Vdash_{\mathfrak{S}} B[v])$,

where u is different from z, $\vec{y}$ and not occurring in A and where v does not appear in $\mathrm{Tr}_{\mathfrak{S}}[p,u]$ nor in $B[x]$.
By $\Vdash_{\mathfrak{S}} A$ we mean $\emptyset\Vdash_{\mathfrak{S}} A$, and if it holds A is said to be *forcible* with $\mathfrak{S}$.

The definition of the forcing relation for connectives (except $\perp$) and quantifiers is a straightforward formalization of Kripke semantics with relation $\subseteq$ and constant domains. Thus we can easily see the following.

Lemma 44. If $A[\vec{x}]$ follows intuitionistically from $B_0[\vec{x}]$, ..., $B_n[\vec{x}]$, then IKP_0^- proves $(p\Vdash_{\mathfrak{S}} B_0[\vec{x}])\to...\to(p\Vdash_{\mathfrak{S}} B_n[\vec{x}])\to(p\Vdash_{\mathfrak{S}} A[\vec{x}])$. Thus, forcibility in WEST_0^{nso} is closed under inferences of intuitionistic logic.

The following is the main result of Section 6 in Sato and Zumbrunnen [37].

Theorem 45. Any axiom (and hence any theorem) of $\mathsf{IKP}_0^\sharp$ as well as (Δ_0^{s-}-MP) is forcible in IKP_0^- with large enough $\mathfrak{S}$. The axiom N-Beta is forcible in IKP_0^- plus N-Beta with large enough $\mathfrak{S}$.
The same statements hold, if we replace the subscript 0 by ω, r or w or even delete it.

More precisely, the forcibility in IKP_0^- follows from Lemmata 65, 66 and 69 of [37], and that in IKP_ω^- and IKP^- follows additionally from Lemma 66 of [37]. The forcibility of Δ_0^{s-} foundation and of the axiom N-Beta follows from the next lemma (the same statement as Lemma 72 in [37]), since Δ_0^{s-} foundation is equivalent to $\forall u,a(x\in u\to\mathrm{Trans}[u]^N\to(\forall y\in u)((\forall z\in y)\neg\neg(z\in a)\to\neg\neg\, y\in a)\to\neg\neg\, x\in a)$ which is in $\mathcal{D}_{res}$.

Definition 46 ($\mathcal{D}_{res}$)**.** The class $\mathcal{D}_{res}$ of $\mathcal{L}_\in$ formulae is the smallest class which contains all strongly negative Δ_0 formulae and which is closed under conjunctions, disjunctions and (universal and existential) quantifiers.

Lemma 47. If A is in $\mathcal{D}_{res}$ then, for large enough $\mathfrak{S}$, $\mathsf{IKP}_0^-\vdash A\leftrightarrow(\Vdash_{\mathfrak{S}} A)$.

Corollary 48. Let A be in $\mathcal{D}_{res}$. If $\mathsf{IKP}_0^\sharp+(\Delta_0^{s-}$-$\mathsf{MP})$ proves A then IKP_0^- proves A. If $\mathsf{IKP}_0^\sharp+(\Delta_0^{s-}$-$\mathsf{MP})$ plus the axiom N-Beta proves A then IKP_0^- plus N-Beta proves A.
The same statements hold, if we replace the subscript 0 by ω, r or w or even delete it.

Remark 49. Strictly speaking, this is not an interpretability but only a local interpretability, for in Lemma 47 $\Vdash_{\mathfrak{S}}$ depends on $\mathfrak{S}$ which must depend on A. Nonetheless, as described in Remarks 23, 46 and 74 of [37], we can make it a non-local interpretability result, by modifying the intermediate system $\mathsf{IKP}_0^\sharp+(\Delta_0^{s-}$-$\mathsf{MP})$ but without affecting the result of the combination with Theorem 52 below, i.e., between IKP_0^- and KP^{int}. However, the combined non-local interpretation is not Π_1-preserving, as mentioned in [37, Remark 76].

6.3. Negative interpretation

The reason why we introduced the semi-constructive set theory is that in the theory we can interpret the axioms of Kripke-Platek set theory with N, except the axiom of extensionality. Actually, we do not need to interpret the axiom of extensionality, since our final goal is to interpret subsystems of second order arithmetic and since the extensionality has no role for this purpose.

Definition 50 (KP^{int})**.** The $\mathcal{L}_\in$-theories KP^{int}, KP_0^{int}, KP_ω^{int}, KP_r^{int} and KP_w^{int} are formulated by the same axioms as IKP^-, IKP_0^-, IKP_ω^-, IKP_r^- and IKP_w^- respectively, but on classical logic with equality.

Since any (Δ_0) formula is equivalent to negative (Δ_0) one over classical logic, the restriction to negative formulae in the formulation of IKP^- does not make sense for KP^{int}.

Definition 51 ($\mathsf{KP}\beta^{int}$)**.** The $\mathcal{L}_\in$-theories $\mathsf{KP}\beta^{int}$, $\mathsf{KP}\beta_0^{int}$, $\mathsf{KP}\beta_\omega^{int}$, $\mathsf{KP}\beta_r^{int}$ and $\mathsf{KP}\beta_w^{int}$ are extensions of KP^{int}, KP_0^{int}, KP_ω^{int}, KP_r^{int} and KP_w^{int} respectively, with Axiom Beta, formulated as follows:

$$\mathrm{Rel}[r] \rightarrow \exists f, b(\mathrm{Fun}[f] \wedge \mathrm{DwCl}[b, a, r] \wedge \mathrm{Prog}[b, a, r] \wedge \mathrm{Clp}[f, b, r]),$$

where the abbreviations are as in Section 3.

The following is the main result of Section 5 in Sato and Zumbrunnen [37].

Theorem 52. The Gödel-Gentzen negative interpretation N interprets all the axioms (and hence theorems) of KP_0^{int} in $\mathsf{IKP}_0^\sharp + (\Delta_0^{s-}\text{-}\mathsf{MP})$ and all the axioms (and hence theorems) of $\mathsf{KP}\beta_0^{int}$ in $\mathsf{IKP}_0^\sharp + (\Delta_0^{s-}\text{-}\mathsf{MP})$ plus the axiom N-Beta.

The same statements hold, if we replace the subscript 0 by ω, r or w or even delete it.

Definition 53 ($\mathcal{C}_{res}$)**.** An $\mathcal{L}_\in$-formula is called weak Σ_1 if it is of the form $\exists y \neg(\forall x \in y)A[x]$ where $A[x]$ is a strongly negative Δ_0 formula without any occurrence of the variable y.

The class $\mathcal{C}_{res}$ of $\mathcal{L}_\in$ formulae is the smallest class which contains all weak Σ_1 formulae and which is closed under conjunctions, implications and universal quantifiers.

Lemma 54. If A is in $\mathcal{C}_{res}$, then there is a strongly negative formula B which is classically equivalent to A such that $\mathsf{IKP}_0^\sharp + (\Delta_0^{s-}\text{-}\mathsf{MP}) \vdash A \leftrightarrow B$.

Corollary 55. Let A be in $\mathcal{C}_{res}$. If KP_0^{int} proves A then $\mathsf{IKP}_0^\sharp + (\Delta_0^{s-}\text{-}\mathsf{MP})$ proves A. If $\mathsf{KP}\beta_0^{int}$ proves A then $\mathsf{IKP}_0^\sharp + (\Delta_0^{s-}\text{-}\mathsf{MP})$ plus the axiom N-Beta proves A.

The same statements hold, if we replace the subscript 0 by ω, r or w or even delete it.

7. Second order arithmetic in intensional set theory

Recall that we consider the language $\mathcal{L}_2$ as a sublanguage of that $\mathcal{L}_\in$ of set theory. The lack of the axiom of extensionality does not affect this. Based on this identification, the following is trivial:

Lemma 56. We have the following containment, via the standard embedding of second order arithmetic into set theory:

(i) $\Sigma_1^1\text{-}\mathsf{AC}_0$ is contained in KP_0^{int};

(ii) $\Sigma_1^1\text{-}\mathsf{AC}$ is contained in KP_ω^{int}.

As for Axiom Beta, we have to check that the absence of extensionality does not affect the well-known containment of $\Sigma_2^1\text{-}\mathsf{AC}_0$ in $\mathsf{KP}\beta$.

Lemma 57. Both $\mathsf{KP}\beta_r^{int}$ and $\mathsf{WEST}_r^{nso}(\mathbb{B})$ prove that $\mathrm{WF}[a, r] \leftrightarrow (\exists f)\mathrm{Clp}[f, a, r]$.

Proof. We are working in $\mathsf{KP}\beta_r^{int}$. The proof in $\mathsf{WEST}_r^{nso}(\mathbb{B})$ is almost same.

First assume $\mathrm{WF}[a, r]$. Then there is f and $b \subseteq a$ such that $\mathrm{Prog}[b, a, r]$ and $\mathrm{Clp}[f, b, r]$. Since $\mathrm{Prog}[b, a, r]$ and $\mathrm{WF}[a, r]$, we have $a \subseteq b$ and hence $\mathrm{Clp}[f, a, r]$.

For the converse, let $\mathrm{Clp}[f, a, r]$ and $\mathrm{Prog}[b, a, r]$. We have to show $a \subseteq b$. We prove

$$(\forall x \in a)(f'x = u \rightarrow x \in b)$$

by $\in$-induction on u. The induction hypothesis is $(\forall y \in a)(f'y = v \rightarrow y \in b)$ for any $v \in u$. Take $x \in a$ with $f'x = u$. To show $x \in b$, because of $\mathrm{Prog}[b, a, r]$, it suffices to show that $y \in b$ for any $y \in a$ with $\langle y, x \rangle \in r$. However, such y satisfies $f'y \in f'x = u$ and hence $y \in b$ by induction hypothesis. Thus, for any $x \in a$, since there is u with $f'x = u$, what we have shown implies $x \in b$. This means $a \subseteq b$. $\square$

Since WF is Π^1_1-complete in second order arithmetic, the Σ_1-ness of this property and Σ_1 collection imply the Π^1_1 axiom of choice in second order arithmetic, which is equivalent to Σ^1_2 axiom of choice.

Corollary 58. We have the following containment, via the standard embedding as in Corollary 56:

(i) Σ^1_2-AC_0 is contained in $\mathsf{KP}\beta^{int}_r$;

(ii) Σ^1_2-AC is contained in $\mathsf{KP}\beta^{int}_w$;

(iii) Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$ is contained in $\mathsf{KP}\beta^{int}$.

The relation between Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$ and KP^{int} is more delicate. First of all, it is known that Σ^1_1 axiom of choice is provable in $\mathsf{ACA}_0 + (\mathsf{BI})$ (see, e.g., [38, Theorem V.8.3, Corollary VII.2.19] or [36, Corollary 2.6]). Second, the existence of a β-model, which is automatically an ω-model of $\mathsf{ACA}_0 + (\mathsf{BI})$ and hence of Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$, is provable in $\mathsf{ACA}_0 + (\Pi^{1-}_1$-$\mathsf{CA})$ (see [38, Lemma VII.2.9]), and thus, by relativizing formulae to such a model, we can construct an interpretation of Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$ in $\mathsf{ACA}_0 + (\Pi^{1-}_1$-$\mathsf{CA})$, where $(\Pi^{1-}_1$-$\mathsf{CA})$ is the comprehension axiom for Π^1_1 formulae without set parameters. Third, $\mathsf{ACA}_0 + (\Pi^{1-}_1$-$\mathsf{CA})$ is reducible, for example by the standard cut-elimination technique, to ID_1. Finally, we can interpret ID_1 in KP^{int}, by interpreting, for a positive operator form $A[x, Z]$, the associated fixed-point predicate $P_A[x]$ as

$$(\exists \alpha \in \mathrm{Ord})(\exists f : \alpha + 1 \to V)\left((\forall \xi \leq \alpha)\left(f(\xi) = \left\{y \in \omega : A\left[y, \bigcup\nolimits_{\eta<\xi} f(\eta)\right]\right\}\right) \wedge x \in f(\alpha)\right).$$

Thus we can conclude that Σ^1_1-$\mathsf{AC}_0 + (\mathsf{BI})$ is reducible to KP^{int} and to the system $\mathsf{AET} + (\mathbb{L}$-$\mathsf{TI})$ of explicit mathematics for arithmetical sentences. However, in this reduction process, cut-elimination is involved and therefore this does not imply the interpretability of Σ^1_1-$\mathsf{AC}_0 + (\mathsf{BI})$ in $\mathsf{AET} + (\mathbb{L}$-$\mathsf{TI})$.

Though we do not know if this reduction can be enhanced to an interpretability, we would like to establish the mutual interpretability between $\mathsf{AET} + (\mathbb{L}$-$\mathsf{TI})$ and KP, because we think that KP is better qualified for a reference system of the strength of generalized predicativity, than Σ^1_1-$\mathsf{AC} + (\mathsf{BI})$. Because of Theorem 12, what remains to do is to construct an interpretation of KP in KP^{int}. This is by so-called bisimulation interpretation $*$ (see e.g., [37, AppendixA.1]).

8. Final Result

Combining the results we have obtained, we reach at the following final result.

Theorem 59. In each pair of the following, (1) the former theory is conservative over the latter for Π^1_1 sentences and (2) the two theories prove the same arithmetical sentences:

(i) Σ^1_1-AC_0 and $\mathsf{AETJ} + (\mathsf{T}$-$\mathsf{I_N})$;

(ii) Σ^1_1-AC and $\mathsf{AETJ} + (\mathbb{L}$-$\mathsf{I_N})$;

(iii) Σ^1_2-AC_0 and $\mathsf{T}_0\!\upharpoonright$;

(iv) Σ^1_2-AC and $\mathsf{T}_0\!\upharpoonright + (\mathbb{L}$-$\mathsf{I_N})$;

(v) Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$ and T_0.

The same results hold between the following pairs:

(ii') KP and $\mathsf{AET} + (\mathbb{L}$-$\mathsf{TI})$;

(v') $\mathsf{KP}\beta$ and T_0.

Proof. (1) Let us consider only the case of T_0, since the other cases can be proved similarly. Let $A[X]$ be an arithmetical formula from $\mathcal{L}_2$. Assume Σ^1_2-$\mathsf{AC} + (\mathsf{BI}) \vdash \forall X A[X]$. Then, by Lemma 58, $\mathsf{KP}\beta^{int} \vdash \forall X A[X]$, with $\forall X A[X]$ being considered as a Π_1 formula $\forall x(x \subseteq \omega \to A[x])$ of $\mathcal{L}_\in$. Since A and A^N are classically equivalent, $\mathsf{KP}\beta^{int} \vdash (\forall x \subseteq \omega)A^N[x]$. Now, by $(\forall x \subseteq \omega)A^N[x] \in \mathcal{C}_{res}$, Corollary 55 implies that $\mathsf{IKP}^\sharp + (\Delta^{s-}_0$-$\mathsf{MP})$ plus the axiom N-Beta proves $(\forall x \subseteq \omega)A^N[x]$. Since $(\forall x \subseteq \omega)A^N[x] \in \mathcal{D}_{res}$, Corollary 48 implies that IKP_0^- plus N-Beta proves $(\forall x \subseteq \omega)A^N[x]$. Because of Lemma 35, Corollary 39 together with Lemma 40 implies that $\mathsf{WEST}^{nso}(\mathbb{B})$ proves the realizability of $(\forall x \subseteq \omega)A^N[x]$. Since $(\forall x \subseteq \omega)A^N[x]$ is negative, Corollary 34 shows that $(\forall x \subseteq \omega)A^N[x]$ is provable in $\mathsf{WEST}^{nso}(\mathbb{B})$. Obviously so is $(\forall x \subseteq \omega)A[x]$, which is identified with $\forall X A[X]$. Finally, Theorem 24 implies $\mathsf{T}_0 \vdash \forall X A[X]$. (2) is from Theorem 11. $\square$

Remark 60. As remarked in Remark 49 these are not interpretability results but only local interpretability results, if we require the preservation for the described classes. However, they are for any subclass Φ that has a universal formula commonly in the senses of Σ^1_1-AC_0 (or KP) and of $\mathsf{AETJ} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})$, because we can take a uniform $\mathfrak{S}$ in Lemma 47. Thus, particularly, we have the Π^0_2-preserving interpretability.

Remark 61. In the cases of (iii)-(v) and (v'), the conservation can actually be enhanced to Π^1_2-sentences: by the Π^1_1-completeness of well-foundedness and by Lemma 57, any Π^1_2 sentence of $\mathcal{L}_2$ has a Π_1 sentence equivalent over both $\mathsf{KP}\beta_r^{int}$ and $\mathsf{WEST}_r^{nso}(\mathbb{B})$; and so the same proof as above shows the Π^1_2 conservation.

Remark 62. As shown in Simpson [38, Exercises VII.3.39], Σ^1_2-AC_0 and $\mathsf{KP}\beta_r$ are mutually interpretable, and hence so are Σ^1_2-AC and $\mathsf{KP}\beta_w$ (as well as Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$ and $\mathsf{KP}\beta$). Thus the same results hold also between $\mathsf{KP}\beta_r$ and $\mathsf{T}_0\!\restriction$ and between $\mathsf{KP}\beta_w$ and $\mathsf{T}_0\!\restriction + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$.

Remark 63. As discussed in Subsection 4.5, we can also have interpretations (i) of Σ^1_1-AC_0 in $\mathsf{BON}(\boldsymbol{\mu})$ plus basic induction on numbers; (ii) of Σ^1_1-AC in $\mathsf{BON}(\boldsymbol{\mu})$ plus full induction on numbers; and establish the mutual interpretability between (iii) KP and $\mathsf{BON}(\boldsymbol{\mu})$ plus transfinite induction along a new predicate **Wf**; (iv) Σ^1_2-AC_0, $\mathsf{KP}\beta_r$ and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus basic induction on numbers; (v) Σ^1_2-AC, $\mathsf{KP}\beta_w$ and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus full induction on numbers; and (vi) Σ^1_2-$\mathsf{AC} + (\mathsf{BI})$, $\mathsf{KP}\beta$ and $\mathsf{BON}(\boldsymbol{\mu}, \boldsymbol{S})$ plus full bar induction.

Acknowledgements

The author would like to express his gratitude to Rico Zumbrunnen, the co-author of the preceding paper [37]. He would also like to thank Solomon Feferman, Gerhard Jäger, Thomas Strahm, Dieter Probst and Tim Rosebrock for their valuable comments, and the two referees for careful reading.

References

[1] Aczel, P., 1978. The type theoretic interpretation of constructive set theory. In: MacIntyre, A., Pacholski, L., Paris, J. (Eds.), Logic Colloquium '77, Amsterdam. Vol. 96 of Stud. Logic Foundations Math. North-Holland, Amsterdam, pp. 55–66.

[2] Aczel, P., 1988. Non-well-founded Sets. Vol. 14 of CSLI Lecture Notes. Stanford University, Center for the Study of Language and Information, Stanford, CA.

[3] Avigad, J., 2000. Interpreting classical theories in constructive ones. J. Symbolic Logic 65 (4), 1785–1812. URL `http://www.andrew.cmu.edu/user/avigad/Papers/interp.pdf`

[4] Barwise, J., 1975. Admissible sets and structures: An approach to definability theory. Perspectives in Mathematical Logic. Springer-Verlag, Berlin. URL `http://projecteuclid.org/euclid.pl/1235418470`

[5] Beeson, M. J., 1985. Foundations of constructive mathematics. Vol. 6 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3). Springer-Verlag, Berlin, metamathematical studies.

[6] Feferman, S., 1975. A language and axioms for explicit mathematics. In: Algebra and logic (Fourteenth Summer Res. Inst., Austral. Math. Soc., Monash Univ., Clayton, 1974). Springer, Berlin, pp. 87–139. Lecture Notes in Math., Vol. 450. URL `http://math.stanford.edu/~feferman/papers/LangAxsExplMaths.pdf`

[7] Feferman, S., 1978. Recursion theory and set theory: a marriage of convenience. In: Fenstad, J. E., Gandy, R. O., Sacks, G. E. (Eds.), Generalized recursion theory, II (Proc. Second Sympos., Univ. Oslo, Oslo, 1977). Vol. 94 of Stud. Logic Foundations Math. North-Holland, Amsterdam, pp. 55–90. URL `http://math.stanford.edu/~feferman/papers/RecSetThMarriage.pdf`

[8] Feferman, S., 1979. Constructive theories of functions and classes. In: Logic Colloquium '78 (Mons, 1978). Vol. 97 of Stud. Logic Foundations Math. North-Holland, Amsterdam, pp. 159–224.
URL `http://math.stanford.edu/~feferman/papers/ConstrFns&Classes.pdf`

[9] Feferman, S., 2001. Notes on operational set theory, I. generalization of "small" large cardinals in classical and admissible set theory, technical notes.
URL `http://math.stanford.edu/~feferman/papers/OperationalST-I.pdf`

[10] Feferman, S., 2009. Operational set theory and small large cardinals. Inform. and Comput. 207 (10), 971–979.
URL `http://math.stanford.edu/~feferman/papers/OST-Final.pdf`

[11] Feferman, S., viewed in 2014. Introduction [to foundations of explicit mathematics (in progress), by Solomon Feferman, Gerhard Jäger, and Thomas Strahm, with the assistance of Ulrik Buchholtz], draft.
URL `http://www.math.stanford.edu/~feferman/papers/DraftIntroFEM.pdf`

[12] Feferman, S., viewed in 2015. The operational perspective: Three routes, draft.
URL `http://www.math.stanford.edu/~feferman/papers/Op_Persp_3Routes.pdf`

[13] Feferman, S., Jäger, G., 1993. Systems of explicit mathematics with non-constructive μ-operator, Part I. Ann. Pure Appl. Logic 65 (3), 243–263.

[14] Glaß, T., Rathjen, M., Schlüter, A., 1997. On the proof-theoretic strength of monotone induction in explicit mathematics. Ann. Pure Appl. Logic 85 (1), 1–46.

[15] Griffor, E., Rathjen, M., 1994. The strength of some martin-löf type theories. Arch. Math. Logic 33, 347–385.
URL `http://www1.maths.leeds.ac.uk/ rathjen/typeOHIO.pdf`

[16] Jäger, G., 1983. A well-ordering proof for Feferman's theory T_0. Arch. Math. Logik Grundlag. 23 (1-2), 65–77.

[17] Jäger, G., 1986. Theories for admissible sets: a unifying approach to proof theory. Vol. 2 of Studies in Proof Theory. Lecture Notes. Bibliopolis, Naples.

[18] Jäger, G., 2005. Metapredicative and explicit Mahlo: a proof-theoretic perspective. In: Logic Colloquium 2000. Logic Colloquium 2000. Assoc. Symbol. Logic, Urbana, IL, pp. 272–293.
URL `http://www.iam.unibe.ch/ltgpub/2005/jae05.pdf`

[19] Jäger, G., Kahle, R., Studer, T., 2001. Universes in explicit methamtics. Ann. Pure Appl. Logic 109 (3), 141–162.
URL `http://www.iam.unibe.ch/ltgpub/2001/jks01.pdf`

[20] Jäger, G., Probst, D., 2011. The Suslin operator in applicative theories: its proof-theoretic analysis via ordinal theories. Ann. Pure Appl. Logic 162 (8), 647–660.
URL `http://www.iam.unibe.ch/ltgpub/2011/jp11.pdf`

[21] Jäger, G., Strahm, T., 2001. Upper bounds for matapredicative Mahlo in explicit mathmatics and admissible set theory. J. Symbolic Logic 66 (2), 935–958.
URL `http://www.iam.unibe.ch/ltgpub/2001/js01.pdf`

[22] Jäger, G., Strahm, T., 2002. The proof-theoretic analysis of the Suslin operator in applicative theories. In: Sieg, W., Sommer, R., Talcott, C. (Eds.), Reflections on the Foundations of Mathematics, Essays in Honor of Solomon Feferman. AK Peters, pp. 270–292.
URL `http://www.iam.unibe.ch/ltgpub/2002/js02a.pdf`

[23] Jäger, G., Strahm, T., 2005. Reflections on reflections in explicit mathematics. Ann. Pure Appl. Logic 136 (1-2), 116–133.
URL `http://www.iam.unibe.ch/ltgpub/2005/js05.pdf`

[24] Jäger, G., Studer, T., 2002. Extending the system T_0 of explicit mathematics: the limit and Mahlo axioms. Ann. Pure Appl. Logic 114 (1–3), 79–101.
URL `http://www.iam.unibe.ch/ltgpub/2002/js02b.pdf`

[25] Jäger, G., Zumbrunnen, R., 2014. Explicit mathematics and operational set theory: some ontological considerations. Bull. Symbolic Logic, in press.
URL `http://www.iam.unibe.ch/ltgpub/2014/jz14.pdf`

[26] Kahle, R., Studer, T., 2000. A theory of explicit mathematics equivalent to ID_1. In: Clote, P., Schwichtenberg, H. (Eds.), Proceedings of Computer Science Logic. Springer, pp. 356–370.
URL `http://www.iam.unibe.ch/ltgpub/2000/ks00.pdf`

[27] Lindström, P., 1997. Aspects of Incompleteness. Vol. 10 of Lecture Notes Logic. Springer-Verlag. URL http://projecteuclid.org/euclid.lnl/1235416274

[28] Marzetta, M., 1993. Universes in the theories of types and names. In: Computer Science Logic (San Miniato, 1992). Vol. 702 of Lecture Notes in Comput. Sci. Springer, Berlin, p. 340–351.

[29] Oosten, J. v., 1997. Extensional realizability. Ann. Pure Appl. Logic 84, 317–349.

[30] Rathjen, M., 1996. Monotone inductive definitions in explicit mathematics. J. Symbolic Logic 61 (1), 125–146. URL http://www1.maths.leeds.ac.uk/~rathjen/mojsl.ps

[31] Rathjen, M., 1998. Explicit mathematics with the monotone fixed point principle. J. Symbolic Logic 63 (2), 509–542. URL http://www1.maths.leeds.ac.uk/~rathjen/UNE.pdf

[32] Rathjen, M., 1999. Explicit mathematics with the monotone fixed point principle. II: Models. J. Symbolic Logic 64 (2), 517–550. URL http://www1.maths.leeds.ac.uk/~rathjen/Muun.pdf

[33] Rathjen, M., 1999. The realm of ordinal analysis. In: Cooper, S. B., Truss, J. K. (Eds.), ets and proofs (Leeds, 1997). Vol. 258 of London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, p. 219–279. URL http://www1.maths.leeds.ac.uk/~rathjen/realm.pdf

[34] Rathjen, M., 2001. Kripke-Platek set theory and the anti-foundation axiom. Math. Log. Q. 47 (4), 435–440. URL http://www1.maths.leeds.ac.uk/~rathjen/kpa.pdf

[35] Rathjen, M., 2002. Explicit mathematics with monotone inductive definitions: a survey. In: Sieg, W., Sommer, R., Talcott, C. (Eds.), Reflections on the Foundationsof Mathematics, Essays in Honor of Solomon Feferman. Lect. Notes Log. AK Peters, pp. 329–346. URL http://www1.maths.leeds.ac.uk/~rathjen/fest.pdf

[36] Rathjen, M., 2014. Constructive Zermelo-Fraenkel set theory and the limited principle of omniscience. Ann. Pure Appl. Logic 165 (2), 563–572. URL http://www1.maths.leeds.ac.uk/~rathjen/CZF_LPO-APAL.pdf

[37] Sato, K., Zumbrunnen, R., 2015. A new model construction by making a detour via intuitionistic theories I: Operational set theory without choice is Π_1-equivalent to KP. Ann. Pure Appl. Logic 166 (2), 121–186. URL http://www.ltg.unibe.ch/node/1643

[38] Simpson, S. G., 2009. Subsystems of Second Order Arithmeitc, 2nd Edition. Perspectives in Logic. Springer, Ithca NY.

[39] Strahm, T., 1999. First steps into metapredicativity in explicit mathematics. In: Sets and proofs (Leeds, 1997). Vol. 258 of London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, p. 383–402. URL http://www.iam.unibe.ch/ltgpub/1999/str99.pdf

[40] Strahm, T., 2002. Wellordering proofs for metapredicative Mahlo. J. Symbolic Logic 67 (1), 260–278. URL http://www.iam.unibe.ch/ltgpub/2002/str02.pdf

[41] Takahashi, S., 1989. Monotone inductive definitions in a constructive theory of functions and classes. Ann. Pure Appl. Logic 42 (3), 255–297.

[42] Troelstra, A., 1977. Aspects of constructive mathematics. In: Barwise, J. (Ed.), Handbook of Mathematical Logic. Vol. 90 of Studies in Logic and the Foundations of Mathematics. North-Holland, Amsterdam-New York-Oxford, pp. 973–1052.

[43] Troelstra, A., 1977. Choice sequences, A chapter of intuitionistic mathematics. Oxford Logic Guides. Clarendon Press, Oxford.

[44] Troelstra, A., 1998. Realizability. In: Buss, S. R. (Ed.), Handbook of Proof Theory. Vol. 137 of Studies in logic and the foundations of mathematics. Elsevier, Amsterdam-New York-Oxford, pp. 407–473.

[45] Tupailo, S., 2003. Realization of constructive set theory into explicit methemtics: a lower bound for impredicative Mahlo universe. Ann. Pure Appl. Logic 120 (1-3), 165–196. URL http://www.iam.unibe.ch/ltgpub/2003/tup03.ps

[46] Tupailo, S., 2004. On the intuitionistic strength of monotone inductive definitions. J. Symbolic Logic 69 (3), 790–798.