Interpreting classical theories
in constructive ones

Jeremy Avigad*

Abstract

A number of classical theories are interpreted in analogous theories that are based on intuitionistic logic. The classical theories considered include subsystems of first- and second-order arithmetic, bounded arithmetic, and admissible set theory.

1   Introduction

Proof theory was developed, in part, as a way to reconcile classical and constructive aspects of mathematical reasoning. Given this historical fact, it is not surprising that over the years proof theorists have invested a good deal of effort in reducing classical theories to constructive ones.

Elegant in its simplicity, the Gödel-Gentzen double-negation interpretation works wonders. For example, it seamlessly reduces classical arithmetic to its intuitionistic counterpart, and with some additional work it can be used to interpret Zermelo-Fraenkel set theory in a variant based on intuitionistic logic [13]. Under the double-negation interpretation, Π_2 sentences of arithmetic, which can be said to carry a theory's "computational" content, are not preserved, but in both these examples a further application of the Friedman-Dragalin interpretation [14] provides an easy and effective way of recovering the Π_2 theorems.

Unfortunately, the double-negation and Friedman-Dragalin translations do not always have desirable effects. For example, the double-negation translation of a classical axiom of choice is no longer an axiom of choice, and when applied to fragments of intuitionistic arithmetic, the Friedman-Dragalin translation increases the complexity of the formulae in the induction axioms. For more general purposes, ordinal analysis and functional interpretation are also powerful tools in the reductionist's kit. But though these approaches yield additional information, from the reductionist point of view the resulting interpretations are somewhat indirect.

In this paper I will discuss a way of "repairing" axioms that are damaged by the double-negation translation, using a method developed by Buchholz [7] and

Coquand and Hofmann [11]. The history is as follows. In [7], Buchholz presented an interpretation of classical theories of iterated inductive definitions, reducing them to intuitionistic versions in which the inductive definitions are given by a strictly positive operator. These can then be further reduced to appropriate Martin-Löf type theories, as in Palmgren [18], or to theories of "accessibility" inductive definitions, also in [7]. Recently, Coquand realized that Buchholz' treatment of a single inductive definition can be adapted to yield a remarkably straightforward reduction of the fragment of arithmetic $I\Sigma_1$ to its intuitionistic version $I\Sigma_1^i$. In [11], Coquand and Hofmann also present an interpretation of Buss' theory of bounded arithmetic S_2^1, that is independent of (and somewhat different from) the one given here.

In Section 2, I will describe a general framework for these interpretations, and in the sections that follow this framework will be instantiated in a number of different settings. In Section 3, I will use it to give a slightly different presentation of Coquand's interpretation, which shows that $I\Sigma_1$ is conservative over its intuitionistic version for Π_2 formulae. I will then adapt the argument to S_2^1, yielding $\forall\Sigma_1^b$ conservation over its intuitionistic counterpart IS_2^1. Since $I\Sigma_1^i$ and IS_2^1 have constructive realizability interpretations, the reductions enable one to extract constructive information from the classical theories as well. The analysis also extends to full classical arithmetic, PA, and the theory S_2.

In Sections 4 and 5, I will consider the Kripke-Platek theory of admissible sets. Here the natural analogue of $I\Sigma_1$ is a variant of KP in which foundation is restricted to formulae that are Σ_1 over the universe of sets. In Section 4, I will use an idea due to Friedman [13] to interpret this theory in an "intensional" version that omits the axiom of extensionality, and, in Section 5, I will apply the framework to interpret the latter in its intuitionistic counterpart.

We will ultimately obtain interpretations for versions of KP with or without the axiom of infinity, and with either full or restricted foundation. The strongest theory analyzed in this way is $KP\omega$, which, by direct interpretation, encompasses a number of important classical theories. These include a theory of arithmetic inductive definitions, ID_1, and a subsystem of second-order arithmetic, $\Pi_1^1\text{-}CA^-$, which allows Π_1^1 comprehension without parameters (for details, see [8]). On the other hand, the intensional intuitionistic version $IKP\omega^{int}$ described below is contained in Aczel's constructive set theory, CZF; the latter can, in turn, be embedded in an appropriate version of Martin-Löf type theory (see [2, 3]). In sum, the analysis below provides a net reduction of a number of classical theories to constructive ones.

In Section 6, I will show that the methods can also be applied in the context of subsystems of second-order arithmetic. In particular, we will see that the theory $\Sigma_1^1\text{-}AC$, which includes arithmetic comprehension and an axiom of choice for arithmetic formulae, is reducible to its intuitionistic counterpart, $\Sigma_1^1\text{-}AC^i$. $\Sigma_1^1\text{-}AC$ is interesting in its own right, but also interprets $\widehat{ID_1}$, which allows a weak form of arithmetic inductive definition. On the other hand, one can interpret $\Sigma_1^1\text{-}AC^i$ in an appropriate version of Martin-Löf type theory,[1] so once

---

[1] More explicitly, Martin-Löf [17] shows that one can interpret $\Sigma_1^1\text{-}AC^i$ in a type theory

2

again we have a net reduction of the classical theory to a constructive one.

I should emphasize that the proof-theoretic equivalences of the classical and intuitionistic theories discussed here are well-known. What is notable about this approach is that it applies uniformly to a wide range of theories, and does not require external "machinery." Because the approach involves effective translations of the classical theories to the constructive ones, we obtain conservation results in which the increase in the lengths of proofs can be bounded by a polynomial; as far as I know, when it comes to S_2^1 and IS_2^1, this sharper form of the conservation result is new.

2 The framework

One thing that makes intuitionistic logic attractive is that the logical connectives have a constructive interpretation, commonly attributed to Brouwer, Heyting, and Kolmogorov. What makes negation, in this context, particularly ornery is that it obliterates any constructive information that a formula might otherwise have held: the interpretation of $\theta \to \bot$ tells us nothing beyond the fact that there cannot possibly be a proof of θ. The remedy offered by the Friedman-Dragalin translation is to insist that $\bot$, along with the other atomic formulae, carry additional information; but the type of information these formulae carry is fixed in advance and remains static throughout the proof. The Buchholz-Coquand methods provide a more dynamic interpretation of $\bot$, by reinterpreting implication as well.

For our purposes it is convenient (but not necessary) to take intuitionistic logic to be given by a system of natural deduction, where derivations yield assertions of the form $\Gamma \Rightarrow \varphi$, i.e. "$\varphi$ follows from the hypotheses in Γ." To describe the method in full generality, let L be any first-order language, and consider the following two-sorted "forcing language" L_f. One sort of L_f has variables corresponding to the universe of L, with the associated constants and functions. The other sort has variables $p, q, r, \ldots$ ranging over "conditions," and there is a binary relation $p \preceq q$ ("p is stronger than q") between objects this sort. Finally, for every n-ary relation symbol $R(x_1, \ldots, x_n)$ of L, there is a corresponding $(n+1)$-ary relation symbol $R'(p, x_1, \ldots, x_n)$ of L_f. Intuitively, $R'(p, x_1, \ldots, x_n)$ asserts that condition p "forces" $R(x_1, \ldots, x_n)$. Then to every formula φ in the language of L, we can inductively associate a formula $p \Vdash \varphi$ in

---

with a single universe, without requiring an elimination rule on the universe. The interpretation is similar to that of the theory $IARI$, described in [15, Section 6]. One obtains $\Sigma_1^1\text{-}AC^i$ from $IARI$ by deleting the inductive generation axiom and substituting arithmetic choice for replacement; the choice schema is interpreted via a straightforward modification of Lemma 6.7 of [15].

the language of L_f, as follows:

$$
\begin{aligned}
p \Vdash R(t_1, \ldots, t_n) &\equiv R'(p, t_1, \ldots, t_n), \text{ for every relation symbol } R \\
p \Vdash (\varphi \wedge \psi) &\equiv p \Vdash \varphi \wedge p \Vdash \psi \\
p \Vdash (\varphi \vee \psi) &\equiv p \Vdash \varphi \vee p \Vdash \psi \\
p \Vdash (\varphi \rightarrow \psi) &\equiv \forall q \preceq p \, (q \Vdash \varphi \rightarrow q \Vdash \psi) \\
p \Vdash \forall x \, \varphi &\equiv \forall x \, p \Vdash \varphi \\
p \Vdash \exists x \, \varphi &\equiv \exists x \, p \Vdash \varphi
\end{aligned}
$$

As usual, $\neg\varphi$ is defined to be $\varphi \rightarrow \bot$, and $\forall q \preceq p \, \theta$ is shorthand for $\forall q \, (q \preceq p \rightarrow \theta)$. $\bot$ is taken to be a 0-ary relation, thereby covered by the first clause. If Γ is a set of formulae $\{\psi_1, \ldots, \psi_k\}$, I will write Γ, φ for $\Gamma \cup \{\varphi\}$, and $p \Vdash \Gamma$ for the set of formulae $\{p \Vdash \psi_1, \ldots, p \Vdash \psi_k\}$. The inscription $\Vdash \varphi$ is read "φ is forced," and means that every condition forces φ.

Proposition 2.1 (substitution) *For every formula φ, variable x, and term t, $p \Vdash \varphi[t/x]$ is given by $(p \Vdash \varphi)[t/x]$.*

Definition 2.2 *A forcing notion is* good *if it satisfies the following conditions:*

1. *$\preceq$ is reflexive and transitive*

2. *The atomic forcing relations are monotone: $p \Vdash \theta$ and $p' \preceq p$ implies $p' \Vdash \theta$, for θ atomic*

3. *For every atomic formula θ, if $p \Vdash \bot$ then $p \Vdash \theta$.*

Notice that these conditions are expressible in L_f.

Proposition 2.3 (monotonicity) *From the assumption that the forcing notion good, in intuitionistic logic one can prove monotonicity for all formulae in L.*

Proof. By induction on formulae η. Monotonicity for atomic formulae takes care of the base case, and the transitivity of $\preceq$ is used when η is of the form $\varphi \rightarrow \psi$. $\square$

Proposition 2.4 *Let Γ be a set of formulae in L, and φ any formula in L. If φ is provable from Γ intuitionistically, then $p \Vdash \varphi$ is provable from $p \Vdash \Gamma$ and the assumption that the forcing notion is good.*

Proof. Use induction on the length of the proof. We can assume that the only rule governing $\bot$ is "ex falso sequitur quodlibet" for atomic formulae; that is, the rule "From $\Gamma \Rightarrow \bot$ conclude $\Gamma \Rightarrow \theta$" for atomic θ. The case in which this is the last inference of the proof is covered by clause 3 in Definition 2.2.

Otherwise, the only interesting cases occur when the last inference is either an introduction or elimination rule for $\rightarrow$. To handle $\rightarrow$-introduction, suppose

the last rule of the proof yields $\Gamma \Rightarrow \varphi \rightarrow \psi$ from $\Gamma, \varphi \Rightarrow \psi$. Assume $p \Vdash \Gamma$, $q \preceq p$, and $q \Vdash \varphi$; we need to show that $q \Vdash \psi$. By monotonicity, we have $q \Vdash \Gamma$; but then $q \Vdash \psi$ follows from the inductive hypothesis.

Dealing with $\rightarrow$-elimination is no more difficult. $\qquad\square$

Notes. The forcing clauses above formalize the usual Kripke semantics, provided $\bot$ is treated as a propositional variable and there is a fixed universe for all the possible worlds. See, for example, [23].

If $\preceq$ is defined from a commutative, idempotent "meet" operation $\wedge$ by

$$q \preceq p \equiv (q \wedge p = q),$$

then the clause for implication is equivalent to

$$\forall q \; (q \Vdash \varphi \rightarrow p \wedge q \Vdash \psi).$$

If $\preceq$ has a greatest element $\emptyset$, then $\Vdash \varphi$ is equivalent to $\emptyset \Vdash \varphi$.

If one focuses one's attention on the negative fragment of intuitionistic logic, which involves only the connectives $\wedge$, $\rightarrow$, and $\forall$, the clauses above are consistent with the intuition that $p \Vdash \varphi$ means, in some sense, that there is a proof of φ from p. We will see in Section 5 that the forcing relation behaves particularly well with respect to these connectives.

The Friedman-Dragalin translation arises from the clauses above in the special case where the partial order is trivial (i.e. has a single element) and there is a fixed formula ψ such that for every atomic formula θ, $p \Vdash \theta$ is just $\theta \vee \psi$.

To apply the framework above one need only find suitable interpretations for the conditions and atomic forcing relations.

Corollary 2.5 *Let T be an intuitionistic theory given by a set of axioms. Suppose one defines a good forcing notion in another theory T', in such a way that T' proves that every axiom of T is forced. Then whenever T proves a formula φ, T' proves that φ is forced.*

Proof. If T proves φ intuitionistically, then there is a set of axioms Γ of T such that $\Gamma \Rightarrow \varphi$ is provable intuitionistically. If p is any condition, the hypothesis and Proposition 2.4 imply that T' proves $p \Vdash \Gamma$ and hence $p \Vdash \varphi$. $\qquad\square$

3 Arithmetic

In this section I will show that $I\Sigma_1$ is conservative over its intuitionistic analogue $I\Sigma_1^i$. The interpretation is essentially that of Coquand and Hofmann [11]; the only difference is that here I will use first-order forcing conditions instead of second-order ones, and divide the interpretation into two steps. In this form, it is easy to extend the results to S_2^1.

$I\Sigma_1$ denotes the fragment of classical arithmetic in which the schema of induction is restricted to Σ_1 formulae. Because any primitive recursive function can be introduced in a definitional extension of $I\Sigma_1$, we can conveniently blur the distinction between this theory and $I\Sigma_1(PRA)$, which has symbols denoting such functions in its language. Identifying primitive recursive relations with their characteristic functions, we can then take the Σ_1 formulae to be of the form $\exists x \, A(x)$, where A is primitive recursive, possibly with free variables other than x. The axioms of $I\Sigma_1$ consist of

1. Quantifier-free defining equations for the primitive recursive functions

2. Σ_1 induction: $\varphi(0) \wedge \forall x \, (\varphi(x) \to \varphi(x')) \to \forall x \, \varphi(x)$, where φ is Σ_1 and x' denotes the successor of x

$I\Sigma_1^i$ denotes the corresponding theory based on intuitionistic logic.

In $I\Sigma_1^i$ one can use induction to prove that equality is decidable, and hence that the law of the excluded middle holds for quantifier-free formulae. Markov's principle for primitive recursive predicates is given by

$$\neg\forall x \, A(x) \to \exists x \, \neg A(x) \qquad (MP_{pr})$$

where A is primitive recursive. The interpretation of $I\Sigma_1$ in $I\Sigma_1^i$ will proceed in two steps: first we will interpret $I\Sigma_1$ in $I\Sigma_1^i + (MP_{pr})$, and then we will interpret the latter theory in $I\Sigma_1^i$.

For the first interpretation, the double-negation interpretation suffices. In this context, the interpretation takes $(\varphi \vee \psi)^N$ to $\neg(\neg\varphi^N \wedge \neg\psi^N)$, takes $(\exists x \, \varphi)^N$ to $\neg\forall x \, \neg\varphi^N$, fixes atomic formulae, and commutes with $\wedge$, $\to$, and $\forall$.

Lemma 3.1 $I\Sigma_1^i + (MP_{pr})$ *proves the double-negation translation of each axiom of* $I\Sigma_1$.

Proof. Using (MP_{pr}) and the law of the excluded middle for atomic formulae, $I\Sigma_1^i$ proves that quantifier-free and Σ_1 formulae are equivalent to their N-translations. As a result, the doubly-negated axioms of $I\Sigma_1$ are equivalent to themselves in $I\Sigma_1^i + (MP_{pr})$. $\qquad\square$

For the remainder of this paper I will say that a formula is "negative" if it is part of the negative fragment of intuitionistic logic, that is, it does not contain any instances of the connectives $\vee$ or $\exists$.[2] Let $\mathcal{C}_{pr}$ denote the set of "almost negative" formulae, that is, the smallest set containing the Σ_1 formulae and closed under conjunction, implication, and universal quantification. Since $I\Sigma_1^i + (MP_{pr})$ proves that any formula in $\mathcal{C}_{pr}$ is equivalent to its N-translation, we have

Theorem 3.2 $I\Sigma_1$ *is conservative over* $I\Sigma_1^i + (MP_{pr})$ *for formulae in* $\mathcal{C}_{pr}$.

---

[2]The word "negative" is, unfortunately, overburdened in the literature; in particular, the usage here differs from that of [7].

Let us now apply the framework of Section 2 to interpret $I\Sigma_1^i + (MP_{pr})$ in $I\Sigma_1^i$. It turns out that the appropriate conditions are finite sets

$$\{\forall x\ A_1(x), \ldots, \forall x\ A_k(x)\} \tag{1}$$

of Π_1 sentences with parameters, i.e. Π_1 formulae together with assignments to their free variables.[3] The ordering $\preceq$ between conditions is defined to be the set containment relation, $\supseteq$. Fixing some reasonable encoding, if $\forall x\ A(x)$ is a Π_1 formula with free variables $\vec{y}$, let $\ulcorner \forall x\ A(x) \urcorner$ denote the function of $\vec{y}$ that returns the code of the corresponding sentence with parameters, and let the variables $p, q, r \ldots$ range over finite (coded) sets of such sentences. I will usually write p, q instead of $p \cup q$ and p, φ instead of $p \cup \{\varphi\}$. Since $q \supseteq p$ is equivalent to $q \cup p = q$, $p \Vdash (\varphi \rightarrow \psi)$ is equivalent to

$$\forall q\ (q \Vdash \varphi \rightarrow p, q \Vdash \psi),$$

which is the characterization I will use to verify the interpretations.

Given any proof in $I\Sigma_1^i + (MP_{pr})$, choose m large enough so that all the primitive recursive relations mentioned have complexity less than m, and let Tr_{pr}^m denote a primitive recursive truth predicate for such relations. (I will come back to this issue below.) If p codes a set of the form (1), define

$$\mathrm{Tr}(p, u) \equiv \bigwedge_{i=1}^{k} A_i(u),$$

where Tr_{pr}^m is used to express the right-hand-side; this asserts that the formulae in p are true at least as far as u is concerned. Define

$$p \vdash \varphi \equiv \exists u\ (\mathrm{Tr}(p, u) \rightarrow \varphi) \tag{2}$$

for arbitrary formulae φ. Intuitively, the witness u in (2) can be interpreted as a "proof" that φ follows from the conjunction of the universal sentences in p, since (2) asserts that φ follows more specifically from their instantiations at u. As one might expect, this "provability" relation is monotone in the first argument: if q and p are conditions such that $q \supseteq p$, then $\mathrm{Tr}(q, u)$ implies $\mathrm{Tr}(p, u)$ and hence $p \vdash \varphi$ implies $q \vdash \varphi$.

Define $p \Vdash \theta$ to be $p \vdash \theta$ when θ is atomic, and extend the forcing relation to arbitrary formulae in the language of arithmetic as in Section 2. It is not difficult to verify that the forcing notion is a good one, according to Definition 2.2. The following lemma shows that when it comes to quantifier-free formulae, the relations $\Vdash$ and $\vdash$ coincide.

Lemma 3.3 *If φ is any quantifier-free formula, then $I\Sigma_1^i$ proves*

$$p \Vdash \varphi \quad \textit{iff} \quad p \vdash \varphi.$$

---

[3]Alternatively one can take these to be Π_1 sentences in an expanded language that has a name for every element of the universe. Of course, in the context of arithmetic these names are not necessary, since every number is denoted by the corresponding numeral; but we will need this more general formulation in Sections 5 and 6 below.

Proof. The proof of this lemma relies heavily on the fact that $I\Sigma_1^i$ proves the law of the excluded middle for quantifier-free formulae. In particular, this implies that for such formulae φ, $p \vdash \varphi$ is equivalent to $\exists u\ (\neg \mathrm{Tr}(p,u) \vee \varphi)$, as well as $\exists u\ \neg \mathrm{Tr}(p,u) \vee \varphi$.

The proof proceeds by induction on the complexity of φ. The cases in which φ is atomic, of the form $\theta \wedge \psi$, or of the form $\theta \vee \psi$ are readily dealt with, using the observations in the preceeding paragraph. When φ is of the form $\theta \to \psi$ we have

$$
\begin{aligned}
p \Vdash (\theta \to \psi) &\leftrightarrow \forall q\ (q \Vdash \theta \to p, q \Vdash \psi) \\
&\leftrightarrow \forall q\ (q \vdash \theta \to p, q \vdash \psi).
\end{aligned}
$$

We need to show that this is equivalent to $p \vdash (\theta \to \psi)$. In the forwards direction, assume the last line of the equivalence holds. Since θ implies $\emptyset \vdash \theta$, we have $\theta \to p \vdash \psi$. This is equivalent to

$$
\neg \theta \vee \exists u\ \neg \mathrm{Tr}(p,u) \vee \psi,
$$

which is equivalent to $p \vdash (\theta \to \psi)$. For the other direction, suppose $p \vdash (\theta \to \psi)$ and $q \vdash \theta$; we need to show $p, q \vdash \psi$. From the assumption, we have

$$
\exists u\ \neg \mathrm{Tr}(p,u) \vee (\theta \to \psi) \quad \text{and} \quad \exists v\ \neg \mathrm{Tr}(q,v) \vee \theta.
$$

Arguing by cases yields
$$
\exists w\ \neg \mathrm{Tr}(p \cup q, w) \vee \psi,
$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 3.4 *For every primitive recursive relation A, $I\Sigma_1^i$ proves the following:*

1. *$\forall x\ A(x) \Vdash \forall x\ A(x)$*

2. *If $p \Vdash \neg \forall x\ A(x)$, then $p \Vdash \exists x\ \neg A(x)$.*

Proof. For 1, we have by definition

$$
\begin{aligned}
\forall x\ A(x) \vdash \forall x\ A(x) \quad &\text{iff} \quad \forall z\ (\forall x\ A(x) \Vdash A(z)) \\
&\text{iff} \quad \forall z\ (\forall x\ A(x) \vdash A(z)) \\
&\text{iff} \quad \forall z\ \exists u\ (\mathrm{Tr}_{pr}^m(\{\ulcorner \forall x\ A(x) \urcorner\}, u) \to A(z)).
\end{aligned}
$$

Given z, simply take $u = z$.

For 2, suppose $p \Vdash \neg \forall x\ A(x)$. Then we have

$$
\forall q\ (q \Vdash \forall x\ A(x) \to p, q \Vdash \perp).
$$

In particular, 1 implies $p, \forall x\ A(x) \Vdash \perp$. In other words, there is an element u such that
$$
\mathrm{Tr}(p,u) \wedge A(u) \to \perp
$$

8

and hence
$$\mathrm{Tr}(p, u) \to \neg A(u).$$

Taking $x = u$ we have

$$\exists x \, \exists u \, (\mathrm{Tr}(p, u) \to \neg A(x)),$$

which is, by definition,

$$\exists x \, p \vdash \neg A(x).$$

By Lemma 3.3 this is equivalent to

$$\exists x \, p \Vdash \neg A(x),$$

which is, by definition, $p \Vdash \exists x \, \neg A(x)$. $\qquad\square$

Lemma 3.5 *$I\Sigma_1^i$ proves that each axiom of $I\Sigma_1^i + (MP_{pr})$ is forced.*

Proof. By Lemma 3.3, a quantifier-free formula is forced iff it is true, so the quantifier-free axioms are reduced to themselves. Lemma 3.4 shows that if p forces the antecedent of (MP_{pr}), then it forces the conclusion as well; so (MP_{pr}) is forced.

Finally, to handle induction, suppose

$$p \Vdash \psi(0) \quad \text{and} \quad p \Vdash \forall x \, (\psi(x) \to \psi(x')).$$

where $\psi(x)$ is Σ_1. The second assumption implies that for every x, if p forces $\psi(x)$, then p forces $\psi(x')$ as well. Since $p \Vdash \psi(x)$ is equivalent to a Σ_1 formula, we can use induction to show that for every x, p forces $\psi(x)$; in other words $p \Vdash \forall x \, \psi(x)$. This shows that the induction axiom for ψ is forced. $\qquad\square$

Let $\mathcal{D}_{pr}$ be the smallest set of formulae containing the quantifier-free ones and closed under conjunction, disjunction, and universal and existential quantification. For formulae φ in $\mathcal{D}_{pr}$, $I\Sigma_1^i$ proves that $\Vdash \varphi$ is equivalent to φ. So we have

Theorem 3.6 *$I\Sigma_1^i + (MP_{pr})$ is conservative over $I\Sigma_1^i$ for formulae in $\mathcal{D}_{pr}$.*

This yields

Theorem 3.7 *$I\Sigma_1$ is conservative over $I\Sigma_1^i$ for Π_2 formulae.*

Proof. Every Π_2 formula is contained in $\mathcal{C}_{pr} \cap \mathcal{D}_{pr}$. $\qquad\square$

Since more general instances of induction translate to instances of induction, we have

Theorem 3.8 *PA and $HA + (MP_{pr})$ are conservative over HA for Π_2 formulae.*

9

There is correspondence between intermediate theories as well. Suppose we start with a Π_n formula $\forall x_1 \, \exists x_2 \, \ldots \, Q x_n \, \varphi$, where $n \geq 2$ and φ is quantifier-free. Then its double-negation translation is intuitionistically equivalent to $\forall x_1 \, \neg \exists x_2 \, \ldots \, Q' x_n \, \neg \varphi^N$, where Q' is $\exists$ if Q is $\forall$, and vice-versa. By definition, the assertion that p forces this latter formula is equivalent to

$$\forall x_1, q \ (q \Vdash \exists x_2 \, \ldots \, Q' x_n \, \neg \varphi^N \to p \cup q \Vdash \bot),$$

and this formula is equivalent (over, say $I\Sigma_1^i$) to one of the form $\forall(\Sigma_n \to \Sigma_1)$. For each $n \geq 2$, C_n be the set of formulas of this form. Since from a classical point of view formulas in C_n are Π_n, and since Π_n and Σ_n induction are equivalent in the classical setting, we have

Theorem 3.9 *For $n \geq 2$, $I\Sigma_n$ is a Π_2 conservative extension of IC_n^i.*

More satisfying results involving the correspondence between classical and intuitionistic fragments of arithmetic are available; see Burr [9].

The argument we have just carried out required a primitive recursive truth predicate Tr_{pr}^m for primitive recursive relations of complexity less than or equal to m. If we measure the complexity of a primitive recursive function by the number of instances of composition and primitive recursion employed in its definition, we can obtain such a predicate as follows. For each m, we define a primitive recursive function $\mathrm{Eval}^m(f, s)$, which evaluates a function of complexity at most m, coded by f, at the list of parameters coded by s. Eval^0 is defined so that it computes the result of applying either a constant, successor, or projection function to its arguments, and then for each m, Eval^{m+1} is defined by cases using Eval^m. (A similar construction can be found in [22, Section 1.5].) We can use Eval^m to define Tr_{pr}^m, and derive the necessary properties in $I\Sigma_1^i$.

Alternatively, we can use a Σ_1 evaluation predicate for the primitive recursive functions (also described in [22]) and take $p \vdash \theta$ to assert that "there is a u and a computation sequence c for $A_1(u), \ldots, A_k(u)$, such that if the computation determines that these are all true, then θ." Choosing this method means that we no longer need to use a separate interpretation for each fixed complexity level m.

For the simplest method of all, note that in any given proof only finitely many relations $A_1, \ldots, A_k$ are mentioned, in which case the "truth predicate" need only work for Π_1 sentences involving these relations. We can represent every such sentence as a pair $\langle i, s \rangle$, where i is a value between 1 and k and s codes an assignment to the free variables of $\forall x \, A_i(x)$. With this representation, Tr_{pr}^m can be defined as a straightforward disjunction of length k, and to carry out the interpretation we only need a minimal theory of finite sets and sequences in $I\Sigma_1^i$.

In the applications that follow, one may be able to use variations of each of these three options. Since the last one requires the least effort and is the most clearly applicable in all cases, let us, for concreteness, adopt this way of interpreting the various forms of Tr^m referred to in the sequel. Though I will neglect to qualify the statements of the lemmata below with the condition that

certain formulae mentioned must have complexity less than m, the reader can readily supply the additional details.

With these considerations in hand the interpretation above can be adapted to Buss' theory of bounded arithmetic S_2^1, and its constructive counterpart, IS_2^1. I will rely on the presentation in Cook and Urquhart [10], which provides a nice account of these two theories and their properties. In fact, it will be more convenient to work with the theories CPV and IPV, which are definitional extensions of S_2^1 and IS_2^1 respectively; these extensions include the terms of Cook's theory PV, which denote polynomial time computable functions. Bounded quantification is defined by

$$\forall x \leq t \; \varphi \equiv \forall x \; (t' \leq x \vee \varphi)$$

and

$$\exists x \leq t \; \varphi \equiv \exists x \; (x \leq t \wedge \varphi).$$

In analogy to the Σ_1 formulae above, a formula is said to be NP if it is of the form $\exists x \leq t \; A(x)$, where A is a PV relation and t is a PV term.

CPV is the classical first-order theory given by

1. Quantifier-free axioms defining the basic symbols of the language

2. NP induction:

$$\varphi(0) \wedge \forall x \; (\varphi(\lfloor x/2 \rfloor) \rightarrow \varphi(x)) \rightarrow \forall x \; \varphi(x)$$

where φ is NP.

IPV is simply CPV based on intuitionistic logic. From [10, Theorem 4.5] we know that IPV proves the law of the excluded middle for quantifier-free formulae (and, in fact, formulae in which every quantifier is *sharply* bounded). In particular, we can take the double-negation translation to fix atomic formulae, and show that the translation of a bounded formula is equivalent to a bounded formula.

A bounded analogue of Markov's principle is given by

$$\neg \forall x \leq t \; A(x) \rightarrow \exists x \leq t \; \neg A(x) \qquad (MP_b)$$

where A is a PV relation.

Finally, in analogy to the almost negative formulae, define $\mathcal{C}_b$ to be the smallest set of formulae containing the NP formulae and closed under conjunction, implication, and universal quantification. The double-negation interpretation yields

Theorem 3.10 CPV *is a conservative extension of* $IPV + (MP_b)$ *for formulae in* $\mathcal{C}_b$.

The interpretation of $IPV + (MP_b)$ in IPV follows the pattern of the interpretation of $I\Sigma_1^i + (MP_{pr})$ in $I\Sigma_1^i$, once we take conditions to be finite sets of sentences of the form $\forall x \leq b \; A(x)$, where $A(x)$ is a relation in PV that may

have parameters, and b is a parameter. Let Tr_b^m be a truth predicate for a sufficiently large set of PV relations with parameters. If p is the set

$$\{\forall x \le b_1\ A_1(x), \ldots, \forall x \le b_k\ A_k(x)\},$$

use Tr_b^m and to define

$$\mathrm{Tr}(p, u) \equiv \bigwedge_{i=1}^{k} (u \le b_i \to A_i(u))$$

and

$$p \vdash \varphi \equiv \exists u \le \max(b_1, \ldots, b_k)\ (\mathrm{Tr}(p, u) \to \varphi)$$

for arbitrary φ. Finally, define $p \Vdash \theta$ to be $p \vdash \theta$ for atomic θ, and then extend the forcing relation to the entire language of IPV.

The following lemma asserts that bounded quantification commutes with the $\Vdash$ operator.

Lemma 3.11 *For every formula φ, IPV proves the following:*

1. $p \Vdash \exists x \le t\ \varphi$ *iff* $\exists x \le t\ p \Vdash \varphi$

2. $p \Vdash \forall x \le t\ \varphi$ *iff* $\forall x \le t\ p \Vdash \varphi$

Proof. As in the proof of Lemma 3.3, if θ is atomic then $p \Vdash \theta$ is equivalent to $\theta \lor p \Vdash \perp$. As a result, we have

$$
\begin{aligned}
p \Vdash \exists x \le t\ \varphi \quad &\text{iff} \quad \exists x\ (p \Vdash x \le t \land p \Vdash \varphi) \\
&\text{iff} \quad \exists x\ ((x \le t \lor p \Vdash \perp) \land p \Vdash \varphi).
\end{aligned}
$$

We need to show that this is equivalent to $\exists x \le t\ p \Vdash \varphi$. One direction is easy; for the other direction, suppose the last line of the equivalence above holds. Given such a value of x, reason by cases: if $x \le t$, we are done. If, on the other hand, $p \Vdash \perp$, then $p \Vdash \varphi[0/x]$, since intuitionistic validities are forced; and again we are done.

The argument for 2 is similar. $\qquad\square$

In analogy with Lemmata 3.3–3.5 we have

Lemma 3.12 *If φ is any quantifier-free formula, then IPV proves*

$$p \Vdash \varphi \quad \text{iff} \quad p \vdash \varphi.$$

Lemma 3.13 *For any PV predicate A, IPV proves the following:*

1. $\forall x \le t\ A(x) \Vdash \forall x \le t\ A(x)$

2. *If $p \Vdash \neg\forall x \le t\ A(x)$, then $p \Vdash \exists x \le t\ \neg A(x)$.*

Lemma 3.14 *IPV proves that each axiom of $IPV + (MP_b)$ is forced.*

Let $\mathcal{D}_b$ be the smallest set of formulae in the language of IPV containing the quantifier-free ones and closed under conjunction, disjunction, and universal and existential quantification.

Theorem 3.15 $IPV + (MP_b)$ *is conservative over* IPV *for formulae in* $\mathcal{D}_b$.

Theorem 3.16 CPV *is conservative over* IPV *for* $\forall NP$ *formulae, that is, formulae of the form* $\forall x\, \exists y \leq t\, A(x,y)$, *where* A *is a PV relation.*

Using the analysis in [10], which shows that CPV and IPV are definitional extensions of S_2^1 and IS_2^1, respectively, this yields

Theorem 3.17 S_2^1 *is conservative over* IS_2^1 *for* $\forall \Sigma_1^b$ *formulae.*

Let S_2 and IS_2 denote the extensions of S_2^1 and IS_2^1 in which induction is allowed for arbitrary bounded formulae. Using Lemma 3.11, we see that Lemma 3.14 still holds for this strengthened form of induction, so we have

Theorem 3.18 S_2 *is conservative over* IS_2 *for* $\forall \Sigma_1^b$ *formulae.*

In [11], Coquand and Hofmann interpret CPV in a second-order version of IPV, and then invoke a result from [10] which reduces this to first-order IPV. Their methods yield a strengthening of Theorem 3.17, where the class of formulas conserved include those of the form $\forall\exists\Sigma_1^b$. On the other hand, because the conservation result from [10] uses a normalization argument, there is the possibility of a superexponential increase in the lengths of proofs. Alternatively, one can derive the stronger conclusion from Theorem 3.17 using Parikh's theorem [19], but once again this allows for superexponential growth. As far as I know, it is still an open question as to whether one can obtain the strengthened version of Theorem 3.17 with a polynomial bound on the increase in the lengths of proofs.

4 Extensionality in admissible set theory

In the context of set theory, the proper analogue of $I\Sigma_1$ is the Kripke-Platek theory of admissible sets, without the axiom of infinity, and with foundation restricted to Σ_1 formulae: in addition to the fact that the two theories can be interpreted in one another, one finds additional structural similarities in the work of Rathjen [20].

I will take the language of set theory to contain only a single binary relation symbol $\in$, and take equality to be *defined* by

$$x = y \equiv \forall z \in x\ (z \in y) \wedge \forall z \in y\ (z \in x).$$

Here bounded quantification is given by

$$\forall y \in x\ \varphi \equiv \forall y\ (y \in x \rightarrow \varphi)$$

and
$$\exists y \in x \; \varphi \equiv \exists y \; (y \in x \wedge \varphi).$$

A formula is said to be Δ_0, or restricted, if every quantifier is bounded; it is said to be Σ_1 (resp. Π_1) if it is either Δ_0 or of the form $\exists y \; \varphi$ (resp. $\forall y \; \varphi$), where φ is Δ_0. The axioms of KP are as follows:

1. Extensionality: $x = y \rightarrow (x \in w \rightarrow y \in w)$

2. Pair: $\exists x \; (y \in x \wedge z \in x)$

3. Union: $\exists x \; \forall z \in y \; \forall w \in z \; (w \in x)$

4. Δ_0 separation: $\exists x \; (\forall z \in x \; (z \in y \wedge \varphi(z)) \wedge \forall z \in y \; (\varphi(z) \rightarrow z \in x))$, where φ is Δ_0 and x does not occur in φ

5. Δ_0 collection: $\forall x \in z \; \exists y \; \varphi(x, y) \rightarrow \exists w \; \forall x \in z \; \exists y \in w \; \varphi(x, y)$, where φ is Δ_0

6. Foundation: $\forall x \; (\forall y \in x \; \psi(y) \rightarrow \psi(x)) \rightarrow \forall x \; \psi(x)$, for arbitrary ψ

I have written Δ_0 separation to emphasize that it is Σ_1; this axiom can be used to derive the more usual forms of pair and union. The foundation axiom as presented here is sometimes also called "set induction," and is equivalent to the assertion that every nonempty definable class of sets has an $\in$-least element. I will use $KP{\upharpoonright}$ to denote the theory in which foundation is restricted to Σ_1 formulae.

The definition of equality given above corresponds to the usual notion of extensional equality between sets. It is easy to prove that this relation is reflexive, symmetric, and transitive; and from axiom 1 we can derive $x = y \rightarrow (\varphi(x) \leftrightarrow \varphi(y))$ for every formula φ. Alternatively, we could have taken equality to be a basic logical symbol having these properties, and then replaced axiom 1 with what was previously the definition. The two approaches are equivalent, and the first is more convenient for our purposes.

We would like to interpret KP and $KP{\upharpoonright}$ in intuitionistic versions. One problem that we will encounter is that extensionality is not well-behaved under the double-negation translation. So let us take KP^{int} and $KP^{int}{\upharpoonright}$ to be "intensional" versions, in which the axiom of extensionality is omitted, and let us consider what life in an intensional universe might be like. One can think of such a universe as consisting of "names" for sets, where, in particular, there may be many names for the empty set; i.e. there may be two sets x and y satisfying $\forall z \; (z \notin x)$ and $\forall z \; (z \notin y)$, while for some w we have $x \in w$ but $y \notin w$. Also, taking $x = \{y, z\}$ to abbreviate

$$y \in x \wedge z \in x \wedge \forall w \in x \; (w = y \vee w = z) \tag{3}$$

is misleading, since it is consistent that $z = \{x, y\}$ and $z' = \{x, y\}$ while $z \neq z'$: z and z' may contain different names for x and y.

Friedman [13] (see also Chapter VIII of [6]) has found an elegant way of interpreting extensionality in an intensional universe: declare all the empty sets

to be "isomorphic" to each other, and, more generally, call two sets isomorphic if (inductively) they have isomorphic elements; then replace elementhood by elementhood up to isomorphism. Here I will show that this approach can be implemented in $KP^{int}\!\upharpoonright$.

A formula is said to be Δ_1 (relative to a theory) if it is provably equivalent to both Σ_1 and Π_1 formulae. The next lemma is standard in admissible set theory, and does not require extensionality.

Lemma 4.1 *$KP^{int}\!\upharpoonright$ proves collection for Σ_1 formulae, and separation for Δ_1 formulae. Also, if $\varphi(R)$ is Δ_0 in a language with a new relation symbol R, and one replaces R with a Δ_1 formula, the result is Δ_1 in $KP^{int}\!\upharpoonright$.*

Proof. One obtains Σ_1 collection by pairing the existentially quantified variables. For Δ_1 separation, note that if $\varphi(y)$ is equivalent to a Σ_1 formula $\exists u\,\psi(y,u)$ as well as to a Π_1 formula $\forall u\,\theta(y,u)$, then (classically) we have $\forall y\,\exists u\,(\psi(y,u) \vee \neg\theta(y,u))$; one can then reduce separation for φ to an instance of Δ_0 separation by first using collection to gather a sufficiently large set of witnesses. The last claim is proved by induction on formulae, again using collection. For details, see [5]. $\square$

Ignoring the caveat above and using $x = \{y,z\}$ to denote (3), I will write "x is an unordered pair" for $\exists y \in x, z \in x\ (x = \{y,z\})$, and $\{y,z\} \in w$ for $\exists x \in w\ (x = \{y,z\})$. We can think of a symmetric relation R as given by a set of unordered pairs, allowing $\{x,x\}$ as a degenerate case. With this in mind, let us write $y \sim_R z$ for $\{y,z\} \in R$, and $y \in \mathit{field}(R)$ for $\exists x \in R\ \exists z \in x\ (x = \{y,z\})$. Call such a relation R an *isomorphism relation* if, for every y and z in the field of R, we have

$$y \sim_R z \leftrightarrow (\forall u \in y\ \exists v \in z\ (u \sim_R v) \wedge \forall v \in z\ \exists u \in y\ (v \sim_R u)). \qquad (4)$$

The definition implies that the field of any isomorphism relation R is transitively closed, i.e. $y \in \mathit{field}(R)$ and $w \in y$ imply $w \in \mathit{field}(R)$. Using foundation with Δ_0 formulae one can also show that any isomorphism relation is an equivalence relation on its field.

The global isomorphism relation we are looking for is given by

$$y \sim z \equiv \exists R\ (\text{``}R\text{ is an isomorphism relation and } y \sim_R z\text{''}).$$

Each of the following four lemmata is provable in $KP^{int}\!\upharpoonright$. The last shows that $y \sim z$ has an equivalent Π_1 definition, and is hence Δ_1.

Lemma 4.2 *$KP^{int}\!\upharpoonright$ proves that for every u and v, there is a set of unordered pairs of elements from u and v; that is, for every u and v,*

$$\exists w\ \forall y \in u, z \in v\ (\{y,z\} \in w).$$

Proof. Fix u and v. For any given y, we have

$$\forall z \in v\ \exists r\ (r = \{y,z\}).$$

15

Use Δ_0 collection to obtain

$$\exists s \; \forall z \in v \; (\{y, z\} \in s).$$

In particular, this is true for every y in u; use Δ_0 collection again to show

$$\exists t \; \forall y \in u \; \exists s \in t \; \forall z \in v \; (\{y, z\} \in s).$$

Apply the union axiom to t to get the desired w. $\qquad \square$

Say that an isomorphism relation is *good* for a set x if every element of x is in the field of R.

Lemma 4.3 *For every set x, there is an isomorphism relation that is good for x.*

Proof. Use Σ_1 induction on x. Suppose the claim is true for every element of x; in other words, for every y in x there is an isomorphism relation R_y good for y. Using collection and union, we can define $\hat{R}$ to be the union of the R_y. Using Lemma 4.2 and Δ_0 separation, let S contain unordered pairs from x satisfying the right side of (4), with $\hat{R}$ in place of R; and let R be $\hat{R} \cup S$. Then R is an isomorphism relation that is good for x. $\qquad \square$

Lemma 4.4 *Suppose R and R' are isomorphism relations. Then for every y and z in both the fields of R and R', $y \sim_R z$ if and only if $y \sim_{R'} z$.*

Proof. Fix R and R'. Since the second sentence of the lemma is equivalent to a Δ_0 assertion in y and z, the result follows from a double induction on y and z, and the definition of an isomorphism relation. $\qquad \square$

Lemma 4.5 *$y \sim z$ is equivalent to the assertion*

$$\forall R \; ((\text{``}R \text{ an isomorphism relation''} \wedge y \in \mathit{field}(R) \wedge z \in \mathit{field}(R)) \to y \sim_R z).$$

Proof. Fix y and z. The pairing axiom and Lemma 4.3 imply that there is an isomorphism relation with y and z in its field, and Lemma 4.4 implies that any two such relations must agree. $\qquad \square$

We have established that $\sim$ is a Δ_1 relation. Using the definition and lemmata above one can show that $\sim$ satisfies

$$y \sim z \leftrightarrow (\forall u \in y \; \exists v \in z \; (u \sim v) \wedge \forall v \in z \; \exists u \in y \; (v \sim u)). \qquad (5)$$

Now define

$$y \in^* w \equiv \exists z \in w \; (y \sim z),$$

and if φ is any formula in the language of set theory, let φ^* denote the formula obtained by replacing $\in$ by $\in^*$. Observe that $(x = y)^*$ is given by

$$\forall z \in^* x \; (z \in^* y) \wedge \forall z \in^* y \; (z \in^* x). \qquad (6)$$

16

Lemma 4.6 *Let φ be any formula. The following are provable in $KP^{int}{\upharpoonright}$:*

 1. $x \in z \rightarrow x \in^* z$

 2. $x \sim y \rightarrow (x \in^* z \rightarrow y \in^* z)$

 3. $x \sim y \leftrightarrow (x = y)^*$

 4. $x \sim y \rightarrow (\varphi^*(x) \leftrightarrow \varphi^*(y))$

 5. $\forall x \in^* z \; \varphi^* \leftrightarrow \forall x \in z \; \varphi^*$

 6. $\exists x \in^* z \; \varphi^* \leftrightarrow \exists x \in z \; \varphi^*$

 7. $x = y \rightarrow (x = y)^*$

Proof. Clause 1 follows from the fact that $\sim$ is reflexive. Clause 2 follows from the definition of $\in^*$, and 3 follows from 2 together with equivalences (5) and (6) above. Clause 4 is proved using induction on φ, with 3 as the base case. The forwards direction of 5 is easy, using 1. For the other direction, suppose $\forall x \in z \; \varphi^*$ and $x \in^* z$. The latter means that there is an $x_0 \in z$ such that $x_0 \sim x$; but then $\varphi^*(x_0)$, and hence $\varphi^*(x)$ by 3. The proof of 6 is similar to that of 5. Clause 7 follows from 5, 1, and the definition of equality. $\qquad\square$

Clauses 5 and 6, together with Lemma 4.1, yield

Corollary 4.7 *The $*$-translation of any Δ_0 formula is Δ_1 in $KP^{int}{\upharpoonright}$.*

Lemma 4.8 *The $*$-translation of each axiom of $KP{\upharpoonright}$ is provable in $KP^{int}{\upharpoonright}$. The same is true for KP and KP^{int}.*

Proof. The fact that the $*$-translation of extensionality is provable in the intensional theory follows from clauses 2 and 3 of Lemma 4.6, and the $*$-translations of pairing and union follow from the corresponding axioms in the intensional theory, using clauses 1 and 5. Translations of instances of Δ_0 separation and collection, as well as Σ_1 or full foundation, are handled using clauses 5 and 6 of Lemma 4.6, Corollary 4.7, and Lemma 4.1. $\qquad\square$

This yields

Theorem 4.9 *If $KP{\upharpoonright}$ proves a formula φ, then $KP^{int}{\upharpoonright}$ proves φ^*; and similarly for KP and KP^{int}.*

In the next section we will consider versions of admissible set theory with an axiom of infinity; we can use Lemma 4.6 to show that Theorem 4.9 still holds with this addition. One may also wish to consider versions of Kripke-Platek set theory in which one has a set $\mathbb{N}$ containing the natural numbers as urelements, as well as the primitive recursive functions and a built-in notion of equality on that set. Once again, Theorem 4.9 still holds for these theories, provided that in the intensional versions we have the usual axioms governing equality on $\mathbb{N}$. The modifications necessary for this interpretation are well described in [13, 6], and pose no additional problems in the present setting.

5   Interpreting intensional KP

Having dealt with extensionality, we can now restrict our attention to the interpretation of KP^{int} and $KP^{int}\upharpoonright$. Let IKP^{int} and $IKP^{int}\upharpoonright$ denote the corresponding theories where the underlying logic is intuitionistic. Our goal is to show that the classical theories are conservative over the intuitionistic ones for a certain class of formulae; for the moment, we will focus our attention on $KP^{int}\upharpoonright$. The argument below is modeled after the one in Section 3, but is more delicate because in $IKP^{int}\upharpoonright$ one can not, in general, prove the law of the excluded middle for Δ_0 formulae. Nonetheless, we will again proceed in two steps, and make use of an intermediate theory based on intuitionistic logic. Many of the lemmata below are patterned after similar ones in [7].

In this setting it turns out that the primitive recursive relations of Section 3 are analogous to *negative* Δ_0 formulae, and the Σ_1 formulae of arithmetic are analogous to what I will call "weak Σ_1" formulae in the language of set theory. These are defined to be formulae of the form

$$\exists w \, \neg\forall x \in w \; \varphi \tag{7}$$

where φ is negative and Δ_0, and w does not appear in φ. Being weak Σ_1 is more restrictive than being Σ_1; formula (7) does not quite assert that there is an x satisfying $\neg\varphi$, but rather that there is a set w of candidates, not all of which satisfy φ.

The intermediate theory $IKP^{int\#}\upharpoonright$ is defined to be the theory based on intuitionistic logic, given by the following axioms:

1. Pair and union: as in KP

2. Δ_0 separation$^\#$: as in KP, except restricted to negative Δ_0 formulae

3. Δ_0 collection$^\#$: $\forall x \in z \; \exists y \; \varphi(x,y) \to \exists w \; \forall x \in z \; \neg\forall y \in w \; \neg\varphi(x,y)$ where φ is Δ_0 and negative

4. Σ_1 foundation$^\#$: as in KP, except restricted to weak Σ_1 formulae

Define the axiom schema

$$\neg\forall x \, \varphi \to \exists w \, \neg\forall x \in w \; \varphi(x) \tag{MP_{res}}$$

where φ is negative and Δ_0. Since the converse direction is intuitionistically valid, (MP_{res}) implies that the negation of any Π_1 formula is equivalent to something that is weak Σ_1.

In the first step we will use the double-negation translation to interpret $KP^{int}\upharpoonright$ in $IKP^{int\#}\upharpoonright + (MP_{res})$. Since we no longer have the decidability of atomic formulae in the latter, here we must take $(y \in x)^N$ to be $\neg\neg y \in x$.

Lemma 5.1 *Let φ be any formula. Then the following are intuitionistically valid:*

1. $y \in x \to \varphi^N$ *iff* $\neg\neg y \in x \to \varphi^N$

2. $(\forall y \in x \ \varphi)^N$ *iff* $\forall y \in x \ \varphi^N$

3. $(\exists y \in x \ \varphi)^N$ *iff* $\neg\forall y \in x \ \neg\varphi^N$

Hence, the double-negation translation of any Δ_0 formula is intuitionistically equivalent to a Δ_0 formula.

Proof. The last claim is proved by induction on formulae, using equivalences 2 and 3.

The right-to-left direction of 1 follows from the fact that $y \in x$ implies $\neg\neg y \in x$ intuitionistically. Conversely, $y \in x \to \varphi^N$ implies $\neg\varphi^N \to \neg y \in x$, and hence $\neg\neg y \in x \to \neg\neg\varphi^N$. But since $\neg\neg\varphi \to \varphi$ is classically valid, $\neg\neg\varphi^N$ implies φ^N intuitionistically.

Regarding 2, we have

$$
\begin{aligned}
(\forall y \in x \ \varphi)^N &\equiv \forall y \ (y \in x \to \varphi)^N \\
&\equiv \forall y \ (\neg\neg y \in x \to \varphi^N)
\end{aligned}
$$

which is equivalent to $\forall y \in x \ \varphi^N$ by part 1. Clause 3 is proved by noting that $\exists y \in x \ \varphi$ and $\neg\forall y \in x \ \neg\varphi$ are classically equivalent, and applying 2 with $\neg\varphi$ in place of φ. $\qquad\square$

Lemma 5.2 *$IKP^{int\#}{\restriction} + (MP_{res})$ proves the double-negation translation of each axiom of $KP^{int}{\restriction}$.*

Proof. Pair and union imply their double-negation translations, and the double-negation translation of any instance of Δ_0 separation is implied by an instance of Δ_0 separation$^\#$.

To handle Δ_0 collection, note that by Lemma 5.1 its double-negation translation is intuitionistically equivalent to

$$\forall x \in z \ \neg\forall y \ \neg\varphi^N(x,y) \to \neg\forall w \ \neg\forall x \in z \ \neg\forall y \in w \ \neg\varphi^N(x,y). \tag{8}$$

Arguing in $IKP^{int\#}{\restriction} + (MP_{res})$, suppose the antecedent is true. By (MP_{res}) we have

$$\forall x \in z \ \exists s \ \neg\forall y \in s \ \neg\varphi^N(x,y).$$

By Δ_0 collection$^\#$ we have

$$\exists w_1 \ \forall x \in z \ \neg\forall s \in w_1 \ \neg\neg\forall y \in s \ \neg\varphi^N(x,y). \tag{9}$$

Given such a set w_1, let w be the set containing $\bigcup w_1$ asserted to exist by the union axiom. Then for every x in z we have

$$\forall y \in w \ \neg\varphi^N(x,y) \to \forall s \in w_1 \ \forall y \in s \ \neg\varphi^N(x,y);$$

weakening the conclusion and taking the contrapositive yields

$$\neg \forall s \in w_1 \, \neg\neg\forall y \in s \, \neg\varphi^N(x,y) \rightarrow \neg\forall y \in w \, \neg\varphi^N(x,y).$$

Combining this with (9) yields

$$\exists w \, \forall x \in z \, \neg\forall y \in w \, \neg\varphi^N(x,y),$$

which implies the conclusion of (8).

Finally, since (MP_{res}) implies that the double-negation of a Σ_1 formula is weak Σ_1, the double-negation of an instance of Σ_1 foundation is equivalent to an instance of Σ_1 foundation$^\#$. $\qquad\square$

Let $\mathcal{C}_{res}$ be the smallest set that is closed under conjunction, implication, and universal quantification, and that contains all weak Σ_1 formulae in which every atomic subformula is preceded by at least one negation. One can show inductively that for each formula φ in $\mathcal{C}_{res}$, $IKP^{int\#}\!\restriction + (MP_{res})$ proves that φ is equivalent to φ^N. Thus we have

Theorem 5.3 $KP^{int}\!\restriction$ is conservative over $IKP^{int\#}\!\restriction + (MP_{res})$ for formulae in $\mathcal{C}_{res}$.

We are now ready to use the forcing framework of Section 2 to reduce $IKP^{int\#}\!\restriction + (MP_{res})$ to $IKP^{int}\!\restriction$. Here the appropriate conditions are finite sets of Π_1 sentences with parameters. Using the third method described in Section 3, let Tr^m_{res} be a Δ_0 truth predicate for sufficiently many Δ_0 sentences with parameters, and if p is a set of Π_1 formulae $\{\forall x \, \varphi_1(x), \ldots, \forall x \, \varphi_k(x)\}$, let

$$\mathrm{Tr}(p,u) \equiv \bigwedge_{i=1}^{k} \forall x \in u \, \varphi_i(x),$$

where Tr^m_{res} is used to express the right-hand side. For any formula φ, define

$$p \vdash \varphi \equiv \exists u \, (\mathrm{Tr}(p,u) \rightarrow \varphi).$$

Intuitively, u provides a "proof" of φ from p by giving a bound on the universal quantifiers that is sufficiently large to witness the fact that φ follows from the formulae in p.

Lemma 5.4 For any formula φ, the following is provable in $IKP^{int}\!\restriction$: If $u \subseteq v$, then

1. $\mathrm{Tr}(p,v)$ implies $\mathrm{Tr}(p,u)$, and

2. $\mathrm{Tr}(p,u) \rightarrow \varphi$ implies $\mathrm{Tr}(p,v) \rightarrow \varphi$.

Lemma 5.5 Let φ and ψ be any formulae, and let θ be any Δ_0 formula. Then the following are provable in $IKP^{int}\!\restriction$:

1. If $p \vdash \varphi$ and $q \supseteq p$ then $q \vdash \varphi$.

2. $\theta \vdash \theta$

3. $p \vdash (\varphi \wedge \psi)$ iff $p \vdash \varphi$ and $p \vdash \psi$.

4. $p \vdash (\theta \to \psi)$ iff $p, \theta \vdash \psi$.

5. If $p \vdash (\varphi \to \psi)$ and $q \vdash \varphi$ then $p, q \vdash \psi$.

6. The following are equivalent:

   (a) $p \vdash \forall x \in z\ \theta$

   (b) $\forall x\ p \vdash (x \in z \to \theta)$

   (c) $\forall x \in z\ p \vdash \theta$.

Proof. Clauses 1 and 2 follow from the definition of $\mathrm{Tr}(p, u)$ and the adequacy of Tr_{res}^m.

For the forward direction of 3, if $\mathrm{Tr}(p, u) \to \varphi \wedge \psi$ then $\mathrm{Tr}(p, u) \to \varphi$ and $\mathrm{Tr}(p, u) \to \psi$. Conversely, if $\mathrm{Tr}(p, u) \to \varphi$ and $\mathrm{Tr}(p, v) \to \psi$ then $\mathrm{Tr}(p, u \cup v) \to \varphi \wedge \psi$ by Lemma 5.4.

For 4, we have

$$\begin{aligned}
\mathrm{Tr}(p, u) \to (\theta \to \psi) \quad &\text{iff} \quad \mathrm{Tr}(p, u) \wedge \theta \to \psi \\
&\text{iff} \quad \mathrm{Tr}(p \cup \{\theta\}, u) \to \psi.
\end{aligned}$$

Regarding 5, if $\mathrm{Tr}(p, u) \to (\varphi \to \psi)$ and $\mathrm{Tr}(q, v) \to \varphi$, then, using the monotonicity of Tr in both arguments, we have $\mathrm{Tr}(p \cup q, u \cup v) \to \psi$.

Finally, regarding 6, (a) implies (b) intuitionistically, and (b) implies (c) intuitionistically as well. To show that (c) implies (a), suppose

$$\forall x \in z\ \exists u\ (\mathrm{Tr}(p, u) \to \theta).$$

Using Δ_0 collection, we have

$$\exists v_1\ \forall x \in z\ \exists u \in v_1\ (\mathrm{Tr}(p, u) \to \theta).$$

Given such a set v_1 we can let v be $\bigcup v_1$, in which case $u \in v_1$ implies $u \subseteq v$; by Lemma 5.4 this shows

$$\exists v\ \forall x \in z\ (\mathrm{Tr}(p, v) \to \theta),$$

which is equivalent to (a). $\qquad\square$

When θ is atomic, define $p \Vdash \theta$ to be $p \vdash \theta$, and extend the forcing relation to arbitrary formulae in the language of set theory according to the clauses in Section 2.

Lemma 5.6 *Let φ be any negative Δ_0 formula. Then in $IKP^{int}{\upharpoonright}$ one can prove*

$$p \Vdash \varphi \quad \text{iff} \quad p \vdash \varphi.$$

Proof. By induction on the complexity of φ. If φ is atomic, there is nothing to do.

If φ is of the form $\theta \wedge \psi$, apply the inductive hypothesis and Lemma 5.5.3. When φ is of the form $\theta \rightarrow \psi$, we have

$$
\begin{aligned}
p \Vdash (\theta \rightarrow \psi) \quad &\text{iff} \quad \forall q \; (q \Vdash \theta \rightarrow p, q \Vdash \psi) \\
&\text{iff} \quad \forall q \; (q \vdash \theta \rightarrow p, q \vdash \psi).
\end{aligned}
$$

We need to show that the last line is equivalent to $p \vdash (\theta \rightarrow \psi)$. In the forwards direction, if we take q to be $\{\theta\}$, we have $p, \theta \vdash \psi$ and hence $p \vdash (\theta \rightarrow \psi)$ by Lemma 5.5.4. Conversely, assuming $p \vdash (\theta \rightarrow \psi)$, Lemma 5.5.5 tells us that if q is any condition and $q \vdash \theta$, then $p, q \vdash \psi$.

Finally, suppose φ is of the form $\forall x \in z \; \theta$. Then we have

$$
\begin{aligned}
p \Vdash \forall x \in z \; \theta \quad &\text{iff} \quad \forall x \; p \Vdash (x \in z \rightarrow \theta) \\
&\text{iff} \quad \forall x \; p \vdash (x \in z \rightarrow \theta).
\end{aligned}
$$

By Lemma 5.5.6 this is equivalent to $p \vdash \forall x \in z \; \theta$. $\qquad\square$

We are almost ready to verify that the axioms of the intermediate theory $IKP^{int\#}{\upharpoonright} + (MP_{res})$ are forced in $IKP^{int}{\upharpoonright}$. The following lemma incorporates many of the technical details.

Lemma 5.7 *Let φ be any negative Δ_0 formula, let be ψ be any weak Σ_1 formula, and let η be arbitrary. Then the following are provable in $IKP^{int}{\upharpoonright}$:*

1. $\Vdash \exists x \; \varphi \;\; \text{iff} \;\; \exists x \; \varphi$

2. $\forall x \; \varphi \Vdash \forall x \; \varphi$

3. *If $p \Vdash \neg \forall x \; \varphi$ then $p \Vdash \exists w \; \neg \forall x \in w \; \varphi$*

4. *If $p \Vdash \forall x \in y \; \eta$ then $\forall x \in y \; p \Vdash \eta$*

5. *If $p \Vdash \forall x \in y \; \exists z \; \varphi$ then $p \Vdash \exists w \; \forall x \in y \; \neg \forall z \in w \; \neg \varphi$*

6. $p \Vdash \forall x \in y \; \psi \;\; \text{iff} \;\; \forall x \in y \; p \Vdash \psi$

Proof. Clause 1 follows from the definition of $\Vdash$ and Lemma 5.6.

Regarding 2, we have

$$
\begin{aligned}
\forall x \; \varphi(x) \Vdash \forall x \; \varphi(x) \quad &\text{iff} \quad \forall z \; (\forall x \; \varphi(x) \Vdash \varphi(z)) \\
&\text{iff} \quad \forall z \; (\forall x \; \varphi(x) \vdash \varphi(z)) \\
&\text{iff} \quad \forall z \; \exists u \; (\mathrm{Tr}^m_{res}(\ulcorner \forall x \in u \; \varphi(x) \urcorner) \rightarrow \varphi(z)).
\end{aligned}
$$

22

Given z, take $u = \{z\}$.

Regarding 3, suppose $p \Vdash \neg\forall x \; \varphi$. Then

$$\forall q \; (q \Vdash \forall x \; \varphi \to p, q \Vdash \bot).$$

In particular, from 2 we have $p, \forall x \; \varphi \Vdash \bot$; that is,

$$\exists u \; (\mathrm{Tr}(p, u) \wedge \forall x \in u \; \varphi \to \bot)$$

and therefore

$$\exists u \; (\mathrm{Tr}(p, u) \to \neg\forall x \in u \; \varphi).$$

Taking $w = u$ we have

$$\exists w \; p \vdash \neg\forall x \in w \; \varphi.$$

By Lemma 5.6 and the definition of $\Vdash$, this is equivalent to

$$p \Vdash \exists w \; \neg\forall x \in w \; \varphi.$$

Regarding 4, suppose $p \Vdash \forall x \in y \; \eta$. Then

$$\forall x, q \; (q \Vdash x \in y \to p, q \Vdash \eta).$$

In particular, if $x \in y$ then $\Vdash x \in y$, and hence $p \Vdash \eta$.

It should be no surprise that proving 5 requires the use of Δ_0 collection. Suppose $p \Vdash \forall x \in z \; \exists y \; \varphi(x, y)$, where φ is Δ_0 and negative; we need to show $p \Vdash \exists w \; \forall x \in z \; \neg\forall y \in w \; \neg\varphi(x, y)$. By 4 and the definition of $\Vdash$, the assumption implies

$$\forall x \in z \; \exists y \; p \Vdash \varphi(x, y)$$

which is equivalent to

$$\forall x \in z \; \exists y \; p \vdash \varphi(x, y)$$

by Lemma 5.6. This is, by definition,

$$\forall x \in z \; \exists y, u \; (\mathrm{Tr}(p, u) \to \varphi(x, y)).$$

Pairing y and u and using Δ_0 collection, we get

$$\exists w, v_1 \; \forall x \in z \; \exists y \in w, u \in v_1 \; (\mathrm{Tr}(p, u) \to \varphi(x, y)).$$

Letting v be $\bigcup v_1$ we have

$$\exists w, v \; \forall x \in z \; \exists y \in w \; (\mathrm{Tr}(p, v) \to \varphi(x, y)),$$

which intuitionistically implies

$$\exists w, v \; \forall x \in z \; (\mathrm{Tr}(p, v) \to \exists y \in w \; \varphi(x, y))$$

and hence

$$\exists w, v \; \forall x \in z \; (\mathrm{Tr}(p, v) \to \neg\forall y \in w \; \neg\varphi(x, y)).$$

23

This last formula is intuitionistically equivalent to

$$\exists w, v \ (\mathrm{Tr}(p, v) \to \forall x \in z \ \neg \forall y \in w \ \neg \varphi(x, y)),$$

which is, by definition,

$$\exists w \ p \vdash \forall x \in z \ \neg \forall y \in w \ \neg \varphi(x, y).$$

By Lemma 5.6 and the definition of $\Vdash$, this is equivalent to

$$p \Vdash \exists w \ \forall x \in z \ \neg \forall y \in w \ \neg \varphi(x, y),$$

as desired.

Finally, regarding 6, note that 4 takes care of the forwards direction. For the converse direction, suppose

$$\forall x \in y \ p \Vdash \exists z \ \neg \forall v \in z \ \varphi$$

where φ is Δ_0 and negative. Then by the definition of $\Vdash$ and Lemma 5.6 we have

$$\forall x \in y \ \exists z \ p \vdash \neg \forall v \in z \ \varphi.$$

Since $p \vdash \neg \forall v \in z \ \varphi$ is Σ_1, we can pair existential quantifiers and use Δ_0 collection to obtain

$$\exists s_1 \ \forall x \in y \ \exists z \in s_1 \ p \vdash \neg \forall v \in z \ \varphi.$$

If s is $\bigcup s_1$, then $z \in s_1$ implies $z \subseteq s$ and so $\neg \forall v \in z \ \varphi$ implies $\neg \forall v \in s \ \varphi$. As a result, the last formula implies

$$\exists s \ \forall x \in y \ p \vdash \neg \forall v \in s \ \varphi.$$

By Lemma 5.5.6, this is equivalent to

$$\exists s \ p \vdash \forall x \in y \ \neg \forall v \in s \ \varphi$$

and hence
$$p \Vdash \exists s \ \forall x \in y \ \neg \forall v \in s \ \varphi.$$

Since intuitionistic validities are forced, this implies

$$p \Vdash \forall x \in y \ \exists s \ \neg \forall v \in s \ \varphi$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 5.8 *$IKP^{int}{\upharpoonright}$ proves that each axiom of $IKP^{int\#}{\upharpoonright} + (MP_{res})$ is forced.*

Proof. If φ is pair, union, or Δ_0 separation$^{\#}$, then by Lemma 5.7.1 $\Vdash \varphi$ is equivalent to φ, which follows from the corresponding axiom of $IKP^{int}\upharpoonright$.

Lemma 5.7.3 implies that (MP_{res}) is forced, and Lemma 5.7.5 takes care of Δ_0 collection$^{\#}$. Finally, to handle Σ_1 foundation$^{\#}$, suppose

$$p \Vdash \forall x \, (\forall y \in x \, \psi(y) \rightarrow \psi(x)),$$

where ψ is weak Σ_1. By definition this means

$$\forall x, q \, (q \Vdash \forall y \in x \, \psi(y) \rightarrow p, q \Vdash \psi(x)).$$

In particular, taking q to be p and applying Lemma 5.7.6, we have that for every x

$$\forall y \in x \, p \Vdash \psi(y) \quad \text{implies} \quad p \Vdash \psi(x).$$

Since $p \Vdash \psi(y)$ is equivalent to a Σ_1 formula, we can use Σ_1 foundation in the target theory to show

$$\forall x \, p \Vdash \psi(x)$$

and hence $p \Vdash \forall x \, \psi(x)$, as desired. $\qquad\square$

Let $\mathcal{D}_{res}$ be the smallest set that contains the negative Δ_0 formulae and is closed under conjunction, disjunction, and universal and existential quantification. For every φ in $\mathcal{D}_{res}$, $IKP^{int}\upharpoonright$ proves that $\Vdash \varphi$ is equivalent to φ. As a result, we have

Theorem 5.9 $IKP^{int\#}\upharpoonright + (MP_{res})$ *is conservative over* $IKP^{int}\upharpoonright$ *for formulae in* $\mathcal{D}_{res}$.

Taken together, Theorems 5.3 and 5.9 yield

Theorem 5.10 *Suppose* $KP^{int}\upharpoonright$ *proves* $\forall x \, \exists y \, \varphi(x, y)$, *where* φ *is* Δ_0. *Then* $IKP^{int}\upharpoonright$ *proves* $\forall x \, \exists w \, \neg\forall y \in w \, \neg\varphi^N(x, y)$.

Proof. Classically the two formulae are equivalent, and the latter is in $\mathcal{C}_{res} \cap \mathcal{D}_{res}$. $\square$

Combining this with Theorem 4.9 yields a reduction of the extensional, classical theory to the intensional, intuitionistic one:

Theorem 5.11 *Suppose* $KP\upharpoonright$ *proves* $\forall x \, \exists y \, \varphi(x, y)$, *where* φ *is* Δ_0. *Then* $IKP^{int}\upharpoonright$ *proves* $\forall x \, \exists w \, \neg\forall y \in w \, \neg\varphi^{*N}(x, y)$.

Proof. By Corollary 4.7, $\varphi^*(x, y)$ is Δ in $KP^{int}\upharpoonright$, and so equivalent to a Σ_1 formula $\exists u \, \theta(x, y, u)$. Pairing quantifiers and applying the previous theorem, we have that $IKP^{int}\upharpoonright$ proves $\forall x \, \exists w \, \neg\forall y \in w, u \in w \, \neg\theta^N(x, y, u)$. The conclusion follows from the fact that $\neg\forall u \in w \, \neg\theta^N(x, y, u)$ implies $\varphi^{*N}(x, y)$. $\qquad\square$

What about adding the full foundation schema to both sides? Interpreting the Σ_1 foundation axiom of $IKP^{int\#}\upharpoonright$ made use of Lemma 5.7.6, which asserts

that $p \Vdash \forall x \in y\ \eta$ is equivalent to $\forall x \in y\ p \Vdash \eta$ when η is weak Σ_1. To interpret KP^{int}, it suffices to allow foundation for negative formulae in $KP^{int\#}$; and to interpret *that*, we need to know that the equivalence given by Lemma 5.7.6 holds for arbitrary negative formulae. This fact is supplied by the following two lemmata.

Lemma 5.12 *Suppose φ and ψ are any two Δ_0 formulae. Then $IKP^{int}\!\upharpoonright$ proves*

$$(\varphi \rightarrow \exists u\ \psi) \quad implies \quad \exists v\ (\varphi \rightarrow \exists u \in v\ \psi)$$

where v is any new variable.

Proof. Arguing in $IKP^{int}\!\upharpoonright$, suppose $\varphi \rightarrow \exists u\ \psi$. We can use pairing and Δ_0 separation to prove the existence of $\{\emptyset\}$, and then the existence of a set z such that

$$\forall w\ (w \in z \leftrightarrow w \in \{\emptyset\} \wedge \varphi).$$

If there is any w in z, then φ holds, and hence so does $\exists u\ \psi$. In other words, we have

$$\forall w \in z\ \exists u\ \psi.$$

By Δ_0 collection we have

$$\exists v\ \forall w \in z\ \exists u \in v\ \psi.$$

But if φ holds then $\emptyset$ is in z, so we have

$$\exists v\ (\varphi \rightarrow \exists u \in v\ \psi)$$

as needed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 5.13 *If φ is negative, $IKP^{int}\!\upharpoonright$ proves*

$$(x \in y \rightarrow p \Vdash \varphi) \quad iff \quad p \Vdash (x \in y \rightarrow \varphi)$$

and hence

$$\forall x \in y\ p \Vdash \varphi \quad iff \quad p \Vdash \forall x \in y\ \varphi.$$

Proof. The second claim follows easily from the first, and proving the right-to-left direction of the first claim is straightforward, as in the proof of Lemma 5.7.4. The left-to-right direction of the first claim is proved by induction on the complexity of φ.

In the atomic case, suppose $x \in y \rightarrow p \Vdash \theta$. Then we have

$$x \in y \rightarrow p \vdash \theta$$

and hence

$$x \in y \rightarrow \exists u\ (\mathrm{Tr}(p, u) \rightarrow \theta).$$

By the previous lemma, there is a v_1 such that

$$x \in y \to \exists u \in v_1 \; (\mathrm{Tr}(p, u) \to \theta).$$

Letting v be $\bigcup v_1$, we have

$$x \in y \to (\mathrm{Tr}(p, v) \to \theta).$$

Rearranging the antecedents in this last formula yields

$$\exists v \; (\mathrm{Tr}(p, v) \to (x \in y \to \theta)),$$

which is the same as

$$p \vdash (x \in y \to \theta).$$

By Lemma 5.6, this is equivalent to $p \Vdash (x \in y \to \theta)$.

Handling the cases involving the connectives $\wedge$, $\to$, and $\forall$ is straightforward, using the following intuitionistic equivalences:

- $(\eta \to (\varphi \wedge \psi)) \leftrightarrow ((\eta \to \varphi) \wedge (\eta \to \psi))$

- $(\eta \to (\varphi \to \psi)) \leftrightarrow ((\eta \to \varphi) \to (\eta \to \psi))$

- $(\eta \to \forall x \; \varphi) \leftrightarrow \forall x \; (\eta \to \varphi), \quad$ if x is not free in η.

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Since we can now interpret foundation for negative formulae as in the proof of Lemma 5.8, we have

Theorem 5.14 *KP is interpretable in IKP^{int}.*

Here and for the rest of this section, I will take the wording of Theorem 5.14 as an abbreviation for the assertion that Theorem 5.11 still holds when one replaces $KP{\upharpoonright}$ and $IKP^{int}{\upharpoonright}$ by the theories mentioned.

We can generalize these results. Suppose φ, ψ, θ, and η are formulae such that

- $KP^{int}{\upharpoonright} + \psi$ proves φ^*,

- $IKP^{int\#}{\upharpoonright} + (MP_{res}) + \theta$ proves ψ^N, and

- $IKP^{int}{\upharpoonright} + \eta$ proves $\Vdash \theta$.

Then $KP{\upharpoonright} + \varphi$ is interpretable in $IKP^{int}{\upharpoonright} + \eta$. For example, suppose we take the axiom of infinity to be given (as in [3]) by

$$\exists x \; (\exists y \in x \; \mathrm{zero}(y) \wedge \forall y \in x \; \exists z \in x \; \mathrm{succ}(y, z)), \qquad\qquad (\textit{infinity})$$

where $\mathrm{zero}(y)$ is the formula $\forall z \in y \; \bot$ and $\mathrm{succ}(y, z)$ is

$$y \in z \wedge \forall u \in y \; (u \in z) \wedge \forall u \in z \; (u \in y \vee u = y).$$

27

Then clauses 1 and 7 of Lemma 4.6 guarantee that $(infinity^*)$ follows from $(infinity)$, and the double-negation translation of $(infinity)$ is intuitionistically implied by

$$\exists x\, (\neg \forall y \in x\, \neg \mathrm{zero}^N(y) \wedge \forall y \in x\, \neg \forall z \in x\, \neg \mathrm{succ}^N(y,z)). \tag{10}$$

Since (10) is in $\mathcal{D}_{res}$, $IKP^{int}{\upharpoonright}$ proves that it is forced if and only if it is true. Furthermore, since (10) is implied by $(infinity)$, we have

Theorem 5.15 $KP\omega{\upharpoonright}$ *is interpretable in* $IKP\omega^{int}{\upharpoonright}$, *and* $KP\omega$ *is interpretable in* $IKP\omega^{int}$.

Here the symbol ω in a theory's name indicates that infinity is to be included among the axioms.

By analogous reasoning we have, for example,

Theorem 5.16 *Let* $\theta(x,y)$ *be any* Δ_0 *formula. Then* $KP\omega + \forall x\, \exists y\, \theta(x,y)$ *is interpretable in* $IKP\omega^{int} + \forall x\, \exists y\, \theta^{*N}(x,y)$.

Suppose that instead of adding an axiom of infinity, we take KPu to be the theory of [16], with a set $\mathbb{N}$ of natural numbers as urelements. Then, according to the discussion at the end of Section 4, KPu is interpreted in $IKPu^{int}$. As a result, if KPu proves $\forall x \in \mathbb{N}\, \exists y \in \mathbb{N}\, A(x,y)$ for some primitive recursive predicate A, then $IKPu^{int}$ proves $\forall x \in \mathbb{N}\, \neg \forall y \in \mathbb{N}\, \neg A(x,y)$. Since the latter theory is closed under the Friedman-Dragalin translation (see [14]), we have

Theorem 5.17 KPu *is conservative over* $IKPu^{int}$ *for* Π_2 *sentences of arithmetic.*

In [16], Jäger also considers theories in which one drops the foundation axiom and replaces it with various forms of induction over the natural numbers. The methods discussed in this section apply to these theories as well: the analogues of Theorem 5.10 hold for the intensional versions of $KPu^0 + (\Sigma_1\ induction)$ and $KPu^0 + (induction)$. However, the corresponding version of Theorem 5.11 does not follow, since the interpretation of extensionality in Section 4 requires Σ_1 foundation.

6 Subsystems of second-order arithmetic

The theory $KPu^0 + (induction)$, discussed at the end of the previous section, has the same strength as the subsystem of second-order arithmetic, $\Sigma_1^1\text{-}AC$ (see [16]). In this section I will show that the methods we have been using can be applied to $\Sigma_1^1\text{-}AC$ directly, reducing it, as well, to its intuitionistic counterpart.[4]

---

[4]The proof-theoretic equivalence of $\Sigma_1^1\text{-}AC$ and $\Sigma_1^1\text{-}AC^i$ is due to Aczel [1]; it can also be obtained via ordinal analysis. The methods described in this section can also be used to interpret the stronger theory $\Sigma_1^1\text{-}DC$ and the weaker theory $\Delta_1^1\text{-}CA$ in their intuitionistic counterparts. Since Friedman [12] has shown that one can interpret both $\Sigma_1^1\text{-}DC$ and $\Sigma_1^1\text{-}AC$ in $\Delta_1^1\text{-}CA$, the analysis here shows that the proof-theoretic strength of all these theories, taken in either the classical or intuitionistic versions, are the same.

Here the task is somewhat easier than that of interpreting KP, for two reasons: we do not have to worry about extensionality, and induction is easier to interpret than foundation.

The language of second-order arithmetic is two-sorted, extending the language of first-order arithmetic with variables $X, Y, Z, \ldots$ ranging over sets of natural numbers, and a relation $\in$ between terms of the two sorts. Equality between second-order objects is taken to be defined in terms of first-order equality, so that $X = Y$ is given by $\forall z \ (z \in X \leftrightarrow z \in Y)$. A formula is said to be *arithmetic* if it contains no second-order quantifiers, though it may contain second-order variables; it is said to be Σ_1^1 if it is either arithmetic or of the form $\exists Y \ \varphi$, where φ is arithmetic.

If we let $\langle \cdot, \cdot \rangle$ denote a primitive recursive pairing function on the natural numbers and read $t \in Y_x$ as $\langle t, x \rangle \in Y$, we can think of the set Y as coding a countable sequence of sets indexed by x. We can also interpret a single set Y as coding a countable *collection* of sets, and introduce bounded second-order quantification by

$$\begin{aligned} \forall X \in Y \ \varphi(X) &\equiv \forall x \ \varphi(Y_x) \\ \exists X \in Y \ \varphi(X) &\equiv \exists x \ \varphi(Y_x) \end{aligned}$$

While this device is suggestive, one should keep in mind that in this context that apparent second-order quantifiers are, in reality, first-order.

The axioms of Σ_1^1-AC are as follows:

1. Quantifier-free defining equations for the first-order symbols of arithmetic

2. Arithmetic comprehension (ACA): $\exists Y \ \forall x \ (x \in Y \leftrightarrow \varphi(x))$, where φ is arithmetic and Y does not appear in φ

3. Arithmetic choice (Σ_1^1-AC): $\forall x \ \exists Y \ \varphi(x, Y) \rightarrow \exists Y \ \forall x \ \varphi(x, Y_x)$, where φ is arithmetic

4. Induction for arbitrary formulae in the language

By coding pairs of sets as a single set, one can easily extend the choice principle to Σ_1^1 formulae, which explains the name. Σ_1^1-AC_0 denotes the theory in which induction is restricted to sets of natural numbers; in the presence of arithmetic comprehension, this set induction axiom implies the schema of induction for arbitrary arithmetic formulae. Σ_1^1-AC^i and Σ_1^1-AC_0^i denote the corresponding intuitionistic theories.

The following lemma makes the analogy to KP more salient.

Lemma 6.1 *Over the other axioms of Σ_1^1-AC_0, (Σ_1^1-AC) is equivalent to*

$$\forall x \ \exists Y \ \varphi(x, Y) \rightarrow \exists W \ \forall x \ \exists Y \in W \ \varphi(x, Y). \qquad (\Sigma_1^1\text{-}AC')$$

Proof. The conclusion of (Σ_1^1-AC) clearly implies the conclusion of (Σ_1^1-AC'): if $\forall x \ \varphi(x, Y_x)$, we can just take W to be Y.

Conversely, suppose $\forall x\ \exists Y \in W\ \varphi(x,Y)$; that is, $\forall x\ \exists y\ \varphi(x,W_y)$. By arithmetic induction, we can show that for every x there is a least number u_x satisfying $\varphi(x,W_{u_x})$. Using arithmetic comprehension to set

$$Y = \{\langle x,v\rangle \mid v \in W_{u_x}\},$$

we have $\forall x\ \varphi(x,Y_x)$, as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

If U is construed as a countable collection of countable collections of sets, let $\bigcup U$ be given by

$$\{\langle x, \langle y,z\rangle\rangle \mid \langle\langle x,y\rangle, z\rangle \in U\}.$$

If we define the subset relation for countable collections of sets by

$$X \sqsubseteq Y \equiv \forall w\ \exists z\ (X_w = Y_z),$$

we then have that for every x, $U_x \sqsubseteq \bigcup U$.

Having seen the developments of the previous section, the reader can anticipate the translation to the current setting. First we need to define an intermediate theory $\Sigma_1^1\text{-}AC^{\#}$, in which negative arithmetic formulae replace the negative Δ_0 formulae of Section 5. The appropriate axioms are:

1. The quantifier-free defining equations for the first-order symbols of arithmetic

2. $(ACA^{\#})$: $\exists Y\ \forall x\ (x \in Y \leftrightarrow \varphi(x))$, where φ is negative and arithmetic and Y does not occur in φ

3. $(\Sigma_1^1\text{-}AC^{\#})$: $\forall x\ \exists Y\ \varphi(x,Y) \rightarrow \exists Y\ \forall x\ \varphi(x,Y_x)$, where φ is negative arithmetic

4. Induction for arbitrary formulae

Let $\Sigma_1^1\text{-}AC_0^{\#} + (\Sigma_1^1\ induction^{\#})$ denote the theory in which induction is restricted to formulae of the form $\exists Y\ \varphi$, where φ is arithmetic and negative. As in Section 5, I will say that a formula is weak Σ_1^1 if it is of the form $\exists W\ \neg\forall X \in W\ \varphi$, where φ is arithmetic and negative and W is not free in φ. Thanks to the fact that induction is easier to interpret than foundation, note that here we do *not* have to restrict Σ_1^1 induction$^{\#}$ to weak Σ_1^1 formulae.

Finally, let us define

$$\neg\forall X\ \varphi \rightarrow \exists W\ \neg\forall X \in W\ \varphi \qquad\qquad (MP_{arith})$$

for φ arithmetic and negative. As expected we have

Lemma 6.2 $\Sigma_1^1\text{-}AC_0^{\#} + (\Sigma_1^1\ induction^{\#})$ *proves the double negation of every axiom of* $\Sigma_1^1\text{-}AC_0 + (\Sigma_1^1\ induction)$*; similarly for* $\Sigma_1^1\text{-}AC^{\#}$ *and* $\Sigma_1^1\text{-}AC$.

Proof. The double-negations of the quantifier-free defining equations are equivalent to themselves, and the double-negation of (ACA) is implied by $(ACA^{\#})$. Using (MP_{arith}), the double-negation of a Σ_1^1 formula is weak Σ_1^1, so each instance of Σ_1^1 induction translates to an instance of Σ_1^1 induction$^{\#}$; and, of course, the arbitrary schema of induction in Σ_1^1-AC translates to induction in Σ_1^1-$AC^{\#}$.

Finally, consider arithmetic choice. By Lemma 6.1, it suffices to interpret the N-translation of $(\Sigma_1^1$-$AC')$. The hypothesis translates to

$$\forall x\ \neg\forall Y\ \neg\varphi^N(x,Y),$$

which, by (MP_{arith}), is equivalent to

$$\forall x\ \exists S\ \neg\forall Y \in S\ \neg\varphi^N(x,Y).$$

Applying $(\Sigma_1^1$-$AC^{\#})$ yields

$$\exists S\ \forall x\ \neg\forall Y \in S_x\ \neg\varphi^N(x,Y).$$

Setting W to $\bigcup S$ we have

$$\exists W\ \forall x\ \neg\forall Y \in W\ \neg\varphi^N(x,Y),$$

which implies the translation of the conclusion of $(\Sigma_1^1$-$AC')$. $\qquad\square$

Let $\mathcal{C}_{arith}$ be the smallest class of formulae that is closed under conjunction, implication, and universal quantification, and that contains all the weak Σ_1^1 formulae in which every subformula of the form $t \in X$ is preceded by at least one negation. Then we have

Theorem 6.3 *Σ_1^1-$AC_0 + (\Sigma_1^1$ induction) is conservative over*

$$\Sigma_1^1\text{-}AC_0^{\#} + (\Sigma_1^1\ induction^{\#}) + (MP_{arith})$$

for formulae in $\mathcal{C}_{res}$, and similarly for Σ_1^1-AC and Σ_1^1-$AC^{\#} + (MP_{arith})$.

To interpret the intermediate theories, let us take our forcing conditions to be finite sets
$$P = \{\forall X\ \varphi_1(X), \dots, \forall X\ \varphi_k(X)\}$$
of Π_1^1 sentences with first- and second-order parameters. In order to code these parameters, conditions must be represented by second-order objects. Let Tr_{arith}^m be a truth predicate for a sufficiently large subset of the arithmetic sentences, and use this to define

$$\mathrm{Tr}(P,U) \equiv \bigwedge_{i=1}^{k} \forall X \in U\ \varphi_i(X).$$

For any formula ψ, define

$$P \vdash \psi \equiv \exists U\ (\mathrm{Tr}(P,U) \to \psi).$$

Then define
$$p \Vdash \theta \equiv p \vdash \theta$$
for atomic formulae θ, and extend the forcing definition to arbitrary formulae in the usual way. Since most of the proofs from Section 5 now carry over, mutatis mutandis, I will only sketch the details below.

Lemma 6.4 *Let φ be any negative arithmetic formula. Then in $\Sigma_1^1\text{-}AC_0^i$ one can prove*
$$P \Vdash \varphi \leftrightarrow P \vdash \varphi.$$

Lemma 6.5 *If φ is any negative arithmetic formula, then the following are provable in $\Sigma_1^1\text{-}AC_0^i$:*

1. $\Vdash \exists X \; \varphi$ iff $\exists X \; \varphi$

2. $\forall X \; \varphi \Vdash \forall X \; \varphi$

3. If $p \Vdash \neg\forall X \; \varphi$ then $p \Vdash \exists W \; \neg\forall X \in W \; \varphi$

Lemma 6.6 *In $\Sigma_1^1\text{-}AC_0^i + (\Sigma_1^1 \; induction)$ one can prove that each axiom of $\Sigma_1^1\text{-}AC_0^\# + (\Sigma_1^1 \; induction^\#) + (MP_{arith})$ is forced; and similarly for $\Sigma_1^1\text{-}AC^i$ and $\Sigma_1^1\text{-}AC^\# + (MP_{arith})$.*

Proof. Lemma 6.5.1 takes care of the quantifier-free axioms and $(ACA^\#)$, and we can verify that induction is forced just as we did for first-order arithmetic, in the proof of Lemma 3.5. Lemma 6.5.3 shows that (MP_{arith}) is forced as well.

Regarding $(\Sigma_1^1\text{-}AC^\#)$, suppose
$$P \Vdash \forall x \; \exists Y \; \varphi(x, Y),$$
where φ is negative and arithmetic. This implies
$$\forall x \; \exists Y \; P \vdash \varphi(x, Y).$$
Since $P \vdash \varphi(x, Y)$ is Σ_1^1, the desired conclusion follows from an application of $(\Sigma_1^1\text{-}AC)$. $\qquad\square$

If $\mathcal{D}_{arith}$ is the smallest set containing the weak Σ_1^1 formulae and closed under conjunction, disjunction, and universal and existential quantification, we have

Theorem 6.7 $\Sigma_1^1\text{-}AC_0^\# + (\Sigma_1^1 \; induction^\#) + (MP_{arith})$ *is conservative over*
$$\Sigma_1^1\text{-}AC_0^i + (\Sigma_1^1 \; induction)$$
for formulae in $\mathcal{D}_{arith}$; and similarly for $\Sigma_1^1\text{-}AC^\# + (MP_{arith})$ and $\Sigma_1^1\text{-}AC^i$.

Combining Theorems 6.3 and 6.7 we have

Theorem 6.8 *If $\Sigma_1^1\text{-}AC_0 + (\Sigma_1^1 \; induction)$ proves $\forall X \; \exists Y \; \varphi(X, Y)$, where φ is arithmetic, then $\Sigma_1^1\text{-}AC_0^i + (\Sigma_1^1 \; induction)$ proves $\exists W \; \forall X \; \exists Y \in W \; \varphi^N(X, Y)$. The corresponding assertion also holds for $\Sigma_1^1\text{-}AC$ and $\Sigma_1^1\text{-}AC^i$.*

Both classical theories are stable under the Friedman-Dragalin translation (see [14, Section 3]). As a result, here too we can recapture the theorems that are arithmetic Π_2.

7 Questions

Given a proof of $\neg\neg\exists x\, A(x)$ in Heyting arithmetic, where A is primitive recursive, one can use either the Friedman-Dragalin translation or the interpretation described above to extract a proof of $\exists x\, A(x)$. What can one can say about the relationship between the two methods?

In a sense, Buchholz' interpretations in [7] are more general than the ones described here, since they allow one to interpret iterations of the basic theory. An "iterated" version of Σ_1^1-AC yields the theory ATR_0, whose main axiom is equivalent to the assertion that every set X is contained in a coded model of Σ_1^1-AC (see [4] or [21]); and the theories KPl and KPi can be seen as "iterated" versions of $KP\omega$, since they axiomatize segments of the constructible set hierarchy that correspond to limits (resp. admissible limits) of admissible ordinals. Can the methods described here be used to provide direct interpretations of ATR_0, KPl, and KPi, in intuitionistic versions thereof? Such an interpretation of KPi would be particularly interesting, because at present the only means of reducing it to its intuitionistic counterpart involves an ordinal analysis.

References

[1] Peter Aczel. The strength of Martin-Löf's intuitionistic type theory with one universe. In Seppo Miettinen and Jouko Väänänen, editors, *Proceedings of the Symposiums of Mathematical Logic, Oulu 1974 and Helsinki 1975*. Dept. of Philosophy, University of Helskinki, 1977.

[2] Peter Aczel. The type theoretic interpretation of constructive set theory. In A. Macintyre, L. Pacholski, and J. Paris, editors, *Logic Colloquium '77*, pages 55–66. North-Holland, Amsterdam, 1978.

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

[4] Jeremy Avigad and Richard Sommer. The model-theoretic ordinal analysis of predicative theories. *Journal of Symbolic Logic*, 64:327–349, 1999.

[5] Jon Barwise. *Admissible Sets and Structures*. Springer, Berlin, 1975.

[6] Michael J. Beeson. *Foundations of Constructive Mathematics*. Springer, Berlin, 1985.

[7] Wilfried Buchholz. The $\Omega_{\mu+1}$-rule. In [8], pages 188–233. 1981.

[8] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*. Lecture Notes in Mathematics #897. Springer, Berlin, 1981.

[9] Wolfgang Burr. Fragments of Heyting-arithmetic. Preprint, 1998.

[10] Stephen A. Cook and Alasdair Urquhart. Functional interpretations of feasibly constructive arithmetic. *Annals of Pure and Applied Logic*, 63:103–200, 1993.

[11] Thierry Coquand and Martin Hofmann. A new method for establishing conservativity of classical systems over their intuitionistic version. *Mathematical Structures in Computer Science*, 9:323–333, 1999.

[12] Harvey M. Friedman. *Subsystems of Set Theory and Analysis*. PhD thesis, Massachusetts Institute of Technology, 1967.

[13] Harvey M. Friedman. The consistency of classical set theory relative to a set theory with intuitionistic logic. *Journal of Symbolic Logic*, 38:315–319, 1973.

[14] Harvey M. Friedman. Classically and intuitionistically provable functions. In H. Müller and D. Scott, editors, *Higher Set Theory*, Lecture Notes in Mathematics #669, pages 21–27. Springer, Berlin, 1978.

[15] E. Griffor and M. Rathjen. The strength of some Martin-Löf type theories. *Archive for Mathematical Logic*, 33:347–385, 1994.

[16] Gerhard Jäger. *Theories for Admissible Sets: A Unifying Approach to Proof Theory*. Bibliopolis, Napoli, 1986.

[17] Per Martin-Löf. An intuitionistic theory of types. Manuscript, 1972. Republished in *Twenty-five years of constructive type theory*, Oxford University, 127–172, 1998.

[18] Erik Palmgren. Type-theoretic interpretation of iterated, strictly positive inductive definitions. *Archive for Mathematical Logic*, 32:75–99, 1992.

[19] Rohit Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, 36:494–508, 1971.

[20] Michael Rathjen. The proof-theoretic characterization of the primitive recursive set functions. *Journal of Symbolic Logic*, 57:954–969, 1992.

[21] Stephen G. Simpson. *Subsystems of Second-Order Arithmetic*. Springer, Berlin, 1998.

[22] A. S. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics #344. Springer, Berlin, 1973.

[23] Dirk van Dalen. *Logic and Structure*. Springer, Berlin, third edition, 1997.