Title
THE STRENGTH OF ABSTRACTION WITH PREDICATIVE COMPREHENSION

Author
WALSH, SEAN

The Strength of Abstraction with Predicative Comprehension

Sean Walsh*

November 13, 2015

Abstract

Frege's theorem says that second-order Peano arithmetic is interpretable in Hume's Principle and full impredicative comprehension. Hume's Principle is one example of an *abstraction principle*, while another paradigmatic example is Basic Law V from Frege's *Grundgesetze.* In this paper we study the strength of abstraction principles in the presence of predicative restrictions on the comprehension schema, and in particular we study a predicative Fregean theory which contains all the abstraction principles whose underlying equivalence relations can be proven to be equivalence relations in a weak background second-order logic. We show that this predicative Fregean theory interprets second-order Peano arithmetic (cf. Theorem 3.2).

Contents

*Department of Logic and Philosophy of Science, 5100 Social Science Plaza, University of California, Irvine, Irvine, CA 92697-5100, U.S.A., swalsh108@gmail.com or walsh108@uci.edu

1

1   Introduction

The main result of this paper is a predicative analogue of Frege's Theorem (cf. Theorem 3.2). Roughly, Frege's theorem says that one can recover all of second-order Peano arithmetic using only the resources of Hume's Principle and second-order logic. This result was adumbrated in Frege's *Grundlagen* of 1884 ([10], [13]) and the contemporary interest in this result is due to Wright's 1983 book *Frege's Conception of Numbers as Objects* ([34]). For more on the history of this theorem, see the careful discussion and references in Heck [20] pp. 4-6 and Beth [1].

More formally, Frege's theorem says that second-order Peano arithmetic is interpretable in second-order logic plus the following axiom, wherein the cardinality operator # is a type-lowering function from second-order entities to first-order entities:

(1.1)   *Hume's Principle* : $\forall\, X, Y\ (\#X = \#Y \leftrightarrow \exists\, \text{bijection}\ f : X \to Y)$

Of course, one theory is said to be *interpretable* in another when the primitives of the interpreted theory can be defined in terms of the resources of the interpreting theory so that the translations of theorems of the interpreted theory are theorems of the interpreting theory (cf. [30] §2 or [22] pp. 96-97 or [17] pp. 148-149 or [27] §2.2). For a proof of Frege's Theorem, see Chapter 4 of Wright's book ([34]) or §2.2 pp. 1688 ff of [29].

The second-order logic used in the traditional proof of Frege's Theorem crucially includes impredicative instances of the comprehension schema. Intuitively, the comprehension schema says that every formula $\varphi(x)$ in one free first-order variable determines a second-order entity:

(1.2) $\exists\, F\ \forall\, x\ (Fx \leftrightarrow \varphi(x))$

The traditional proof of Frege's Theorem uses instances of this comprehension schema in which some of the formulas in question contain higher-order quantifiers (cf. [29] p. 1690 equations (44)-(45)). However, there is a long tradition of *predicative mathematics*, in which one attempts to ascertain how much one can accomplish without directly appealing to such instances of the comprehension schema. This was the perspective of Weyl's great book *Das Kontinuum* ([33]) and has been further developed in the work of Feferman ([7], [8]). Many of us today learn and know of this tradition due to its close relation to the system $\mathtt{ACA}_0$ of Friedman and Simpson's project of reverse mathematics ([15], [26]).

However, outside of the inherent interest in predicative mathematics, considerations related to Frege's philosophy of mathematics likewise suggest adopting the predicative perspective. For, Wright and Hale ([18], cf. [4]) have emphasized that Hume's Principle (1.1) is a special instance of the following:

(1.3)   *A[E]* : $\forall\, X, Y\ (\partial_E(X) = \partial_E(Y) \leftrightarrow E(X,Y))$

wherein $E(X,Y)$ is a formula of second-order logic and ∂_E is a type-lowering operator taking second-order entities and returning first-order entities. These principles were called *abstraction principles* by Wright and Hale, who pointed out that the following crucial fifth axiom of Frege's *Grundgesetze* of 1893 and 1903 ([11], [14]) was also an abstraction principle:

(1.4)   *Basic Law V* : $\forall\, X, Y\ (\partial(X) = \partial(Y) \leftrightarrow X = Y)$

2

The operator ∂ as governed by Basic Law V is called the *extension* operator and the first-order entities in its range are called *extensions*. Regrettably, there is no standard notation for the extension operator, and so some authors write $\S X$ in lieu of $\partial(X)$. In what follows, the symbol ∂ without any subscripts will be reserved for the extension operator, whereas the subscripted symbols ∂_E will serve as the notation for the type-lowering operators present in arbitrary abstraction principles (1.3).

While the Russell paradox shows that Basic Law V is inconsistent with the full comprehension schema (1.2) (cf. [29] p. 1682), nevertheless Basic Law V is consistent with predicative restrictions, as was shown by Parsons ([25]), Heck ([19]), and Ferreira-Wehmeier ([9]). This thus suggests the project of understanding whether there is a version of Frege's theorem centered around the consistent predicative fragments of the *Grundgesetze*. This project has been pursued in the last decades by many authors such as Heck ([19]), Ganea ([16]), and Visser ([28]). Their results concerned the restriction of the comprehension schema (1.2) to the case where no higher-order quantifiers are permitted. One result from this body of work says that Basic Law V (1.4) coupled with this restriction on the comprehension schema is mutually interpretable with Robinson's Q. Roughly, Robinson's Q is the fragment of first-order Peano arithmetic obtained by removing all the induction axioms. (For a precise definition of Robinson's Q, see [17] p. 28, [26] p. 4, [29] p. 1680, [30] p. 106). Additional work by Visser allows for further rounds of comprehension and results in systems mutually interpretable with Robinson's Q plus iterations of the consistency statement for this theory, which are likewise known to be interpretable in other weak arithmetics ([28] p. 147). In his 2005 book ([3]), Burgess surveys these kinds of developments, and writes:

> [...] I believe that no one working in the area seriously expects to get very much further in the sequence Q_m while working in predicative Fregean theories of whatever kind ([3] p. 145).

Here Q_m is the expansion of Robinson's Q by finitely many primitive recursive function symbols and their defining equations along with induction for bounded formulas ([3] pp. 60-63), so that Burgess records the prediction that predicative Fregean theories will be interpretable in weak arithmetics.

The main result of this paper suggests that this prediction was wrong, and that predicative Fregean theories can interpret strong theories of arithmetic (cf. Theorem 3.2). While we turn presently to developing the definitions needed to precisely state this result, let us say by way of anticipation that part of the idea is to work both with (i) an expanded notion of a "Fregean theory," so that it includes several abstraction principles, such as Basic Law V, in addition to Hume's Principle, and (ii) an expanded notion of "predicativity," in which one allows some controlled instances of higher-order quantifiers within the comprehension schema (1.2). Hence, of course, it might be that Burgess and others had merely conjectured that predicative Fregean theories in a more limited sense were comparatively weak.

This paper is part of a series of three papers, the other two being [31] and [32]. These papers collectively constitute a sequel to our paper [29], particularly as it concerns the methods and components related to Basic Law V. In that earlier paper, we showed that Hume's Principle (1.1) with predicative comprehension did not interpret second-order Peano

arithmetic with predicative comprehension (cf. [29] p. 1704). Hence at the outset of that paper, we said that "in this specific sense there is no predicative version of Frege's Theorem" ([29] p. 1679). The main result of this present paper (cf. Theorem 3.2) is that when we enlarge the theory to a more inclusive class of abstraction principles containing Basic Law V, we do in fact succeed in recovering arithmetic.

This paper depends on [31] only in that the consistency of the predicative Fregean theory which we study here was established in that earlier paper (cf. discussion at close of next section). In the paper [32], we focus on embedding the system of the *Grundgesetze* into a system of intensional logic. The alternative perspective of [32] then suggests viewing the consistent fragments of the *Grundgesetze* as a species of intensional logic, as opposed to an instance of an abstraction principle.

This paper is organized as follows. In §2 we set out the definitions of the predicative Fregean theory. In §3 it is shown how this predicative Fregean theory can recover full second-order Peano arithmetic. In §4 it is noted that some theories which are conceptually proximate to the predicative Fregean theory are nonetheless inconsistent.

2 Defining a theory of abstraction with predicative comprehension

The predicative Fregean theory with which we work in this paper is developed within the framework of second-order logic. The language L_0 of the background second-order logic is an ω-sorted system with sorts for first-order entities, unary second-order entities, binary second-order entities etc. Further, following the Fregean tradition, the first-order entities are called *objects*, the unary second-order entities are called *concepts*, and the n-ary second-order entities for $n \geq 1$ are called *n-ary concepts*. Rather than introduce any primitive notation for the different sorts, we rather employ the convention of using distinctive variables for each sort: objects are written with lower-case Roman letters $x, y, z, a, b, c \ldots$, concepts are written with upper-case Roman letters $X, Y, Z, A, B, C, F, G, H, U, \ldots$, n-ary concepts for $n > 1$ are written with the upper case Roman letters R, S, T, and n-ary concepts are written with the Roman letters f, g, h when they are graphs of functions.

Besides the sorts, the other basic primitive of the signature of the background second-order logic L_0 are the predication relations. One writes Xa to indicate that object a has property or concept X. Likewise, there are predication relations for n-ary concepts, which we write as $R(a_1, \ldots, a_n)$. The final element of the signature L_0 of the background second-order logic are the projection symbols. The basic idea is that one wants, primitive in the signature L_0, a way to move from the binary concept R and the object a to its projection $R[a] = \{b : R(a, b)\}$. We assume that the signature L_0 of the background second-order logic is equipped with symbols $(R, a_1, \ldots, a_m) \mapsto R[a_1, \ldots, a_m]$ from $(m+n)$-ary concepts R and an m-tuple of objects $(a_1, \ldots, a_m)$ to an n-ary concept $R[a_1, \ldots, a_m] = \{(b_1, \ldots, b_n) : R(a_1, \ldots, a_m, b_1, \ldots, b_n)\}$. Further, typically in what follows we avail ourselves of the tuple notation $\bar{a} = a_1, \ldots, a_n$ and thus write predication and projection more succinctly as $R(\bar{a})$

and $R[\overline{a}]$, respectively.

All this in place, we can then formally define the signature L_0 of the background second-order logic as follows:

Definition 2.1. *The signature L_0 of the background second-order logic is a many-sorted signature which contains (i) a sort for objects and for each $n \geq 1$ a sort for n-ary concepts, (ii) for each $n \geq 1$, an $(n+1)$-ary predication relation symbol $R(a_1, \ldots, a_n)$ which holds between an n-ary concept R and an n-tuple of objects $a_1, \ldots, a_n$, and (iii) for each $n, m \geq 1$, an $(m+1)$-ary projection function symbol $(R, a_1, \ldots, a_m) \mapsto R[a_1, \ldots, a_m]$ from an $(m+n)$-ary concept R and an m-tuple of objects $(a_1, \ldots, a_m)$ to an n-ary concept $R[a_1, \ldots, a_m]$.*

As is usual in many-sorted signatures, we adopt the convention that each sort has its own identity symbol, so that technically cross-sortal identities are not well-formed. But we continue to write all identities with the usual symbol "=" for the ease of readability.

The expansions of second-order logic with which we work are designed to handle abstraction principles (1.3). Hence, suppose that L is an expansion of L_0. Suppose that $E(R, S)$ is an L-formula with two free n-ary relation variables for some $n \geq 1$, with all free variables of $E(R, S)$ explicitly displayed. Then we may expand L to a signature $L[\partial_E]$ which contains a new function symbol ∂_E which takes n-ary concepts R and returns the object $\partial_E(R)$. Then the following axiom, called *the abstraction principle associated to E*, is an $L[\partial_E]$-sentence:

$$(2.1) \qquad A[E]: \quad \forall\, R, S\ (\partial_E(R) = \partial_E(S) \leftrightarrow E(R, S))$$

This generalizes the notion of an abstraction principle (1.3) described in the previous section in that the domain of the operator ∂_E can be n-ary concepts for any specific $n \geq 1$.

This generalization is warranted by several key examples, such as that of ordinals. Let R be a binary concept and let $\mathrm{Field}(R)$ be the unary concept F such that Fx iff there is a y such that Rxy or Ryx. Then consider the following formula $E(R, S)$ on binary concepts:

$$(2.2) \qquad [(\mathrm{Field}(R), R) \models \mathrm{wo} \vee (\mathrm{Field}(S), S) \models \mathrm{wo}] \rightarrow$$
$$\exists\ \text{isomorphism } f : (\mathrm{Field}(R), R) \rightarrow (\mathrm{Field}(S), S)$$

In this, "wo" denotes the natural sentence in the signature of second-order logic which says that a binary concept is a well-order, i.e. a linear order such that every non-empty subconcept of its domain has a least element. It's not too difficult to see that $E(R, S)$ is an equivalence relation on binary concepts, and that two well-orders will be E-equivalent if and only if they are order-isomorphic. Just as the Russell paradox shows that Basic Law V (1.4) is inconsistent with the full comprehension schema, so one can use the Burali-Forti paradox to show that $A[E]$ for this E in equation (2.2) is inconsistent with the full comprehension schema (cf. [21] p. 138 footnote, [2] pp. 214, 311). To handle these abstraction principles we need to adopt restrictions on the comprehension schema, to which we presently turn.

There are three traditional predicative varieties of the comprehension schema: the first-order comprehension schema, the Δ_1^1-comprehension schema, and the Σ_1^1-choice schema (cf. [26] VII.5-6, [29] Definition 5 p. 1683). However, to make the comparison with the full comprehension schema (1.2) precise, we should restate it to include not only concepts but n-ary concepts for all $n \geq 1$ and to indicate its explicit dependence on a signature:

Definition 2.2. *Suppose that L is an expansion of L_0. Then the* Full Comprehension Schema *for L-formulas consists of all axioms of the form $\exists\, R\, \forall\, \overline{a}\, (R\overline{a} \leftrightarrow \varphi(\overline{a}))$, wherein $\varphi(\overline{x})$ is an L-formula, perhaps with parameters, and $\overline{x}$ abbreviates $(x_1, \dots, x_n)$ and R is an n-ary concept variable for $n \geq 1$ that does not appear free in $\varphi(\overline{x})$.*

The most restrictive predicative version of the comprehension schema is then the following, where the idea is that *no* higher-order quantifiers are allowed in the formulas:

Definition 2.3. *Suppose that L is an expansion of L_0. The* First-Order Comprehension Schema *for L-formulas consists of all axioms of the form $\exists\, R\, \forall\, \overline{a}\, (R\overline{a} \leftrightarrow \varphi(\overline{a}))$, wherein $\varphi(\overline{x})$ is an L-formula with no second-order quantifiers but perhaps with parameters, and $\overline{x}$ abbreviates $(x_1, \dots, x_n)$ and R is an n-ary concept variable for $n \geq 1$ that does not appear free in $\varphi(\overline{x})$.*

A more liberal version of the comprehension schema is the so-called Δ_1^1-comprehension schema. A Σ_1^1-formula (resp. Π_1^1-formula) is one which begins with a block of existential quantifiers (resp. universal quantifiers) over n-ary concepts for various $n \geq 1$ and which contains no further second-order quantifiers. One then defines:

Definition 2.4. *Suppose that L is an expansion of L_0. Then the Δ_1^1-Comprehension Schema for L-formulas consists of all axioms of the form*

$$(2.3) \qquad (\forall\, \overline{x}\, \varphi(\overline{x}) \leftrightarrow \psi(\overline{x})) \to \exists\, R\, \forall\, \overline{a}\, (R\overline{a} \leftrightarrow \varphi(\overline{a}))$$

wherein $\varphi(\overline{x})$ is a Σ_1^1-formula in the signature of L and $\psi(\overline{x})$ is a Π_1^1-formula in the signature of L that may contain parameters, and $\overline{x}$ abbreviates $(x_1, \dots, x_n)$, and R is an n-ary concept variable for $n \geq 1$ that does not appear free in $\varphi(\overline{x})$ or $\psi(\overline{x})$.

Finally, traditionally one also includes amongst the predicative systems the following choice principle:

Definition 2.5. *Suppose that L is an expansion of L_0. The Σ_1^1-Choice Schema for L-formulas consists of all axioms of the form*

$$(2.4) \qquad [\forall\, \overline{x}\, \exists\, R'\, \varphi(R', \overline{x})] \to \exists\, R\, [\forall\, \overline{x}\, \varphi(R[\overline{x}], \overline{x})]$$

wherein the L-formula $\varphi(R', \overline{x})$ is Σ_1^1, perhaps with parameters, and $\overline{x}$ abbreviates $(x_1, \dots, x_m)$ and R is an $(m + n)$-ary concept variable for $n, m \geq 1$ that does not appear free in $\varphi(R', \overline{x})$ where R' is an n-ary concept variable.

The Σ_1^1-Choice Schema and the First-Order Comprehension Schema together imply the Δ_1^1-Comprehension Schema (cf. [26] Theorem V.8.3 pp. 205-206, [29] Proposition 6 p. 1683). Hence, even if one's primary interest is in the latter schema, typically theories are axiomatized with the two former schemas since they are deductively stronger, and that is how we proceed in this paper.

To the signature L_0 of the weak background second-order logic, we want to associate a certain weak background L_0-theory. Some of the axioms of this background theory axiomatize the behavior of the predication symbols and the projection symbols. For each $m \geq 1$, one has the following *extensionality axiom*, wherein R, S are m-ary concept variables and $\bar{a} = a_1, \ldots, a_m$ are object variables:

(2.5) $$\forall\, R, S\, [R = S \leftrightarrow (\forall \bar{a}\, (R(\bar{a}) \leftrightarrow S(\bar{a})))]$$

But it should be noted that some authors don't explicitly include the identity symbol for concepts or higher-order entities and simply take it as an abbreviation for coextensionality (cf. [26] pp. 2-3, [3] pp. 14-15). Second, for each $n, m \geq 1$, one has the following *projection axioms* governing the behavior of the projection symbols, wherein R is an $(m+n)$-ary concept variable and $\bar{a} = a_1, \ldots, a_m, \bar{b} = b_1, \ldots, b_n$ are object variables:

(2.6) $$\forall\, R\, \forall\, \bar{a}, \bar{b}\, [(R[\bar{a}])(\bar{b}) \leftrightarrow R(\bar{a}, \bar{b})]$$

Finally, with all this in place, we can define the weak background theory of second-order logic:

Definition 2.6. *The weak background theory of second-order logic Σ_1^1-$\mathtt{OS}$ is L_0-theory consisting of (i) the extensionality axioms (2.5) and the projection axioms (2.6) and (ii) the Σ_1^1-Choice Schema for L_0-formulas (Definition 2.5) and (iii) the First-Order Comprehension Schema for L_0-formulas (Definition 2.3).*

In the theory Σ_1^1-$\mathtt{OS}$ and its extensions, we use standard abbreviations for various operations on concepts, for instance $X \cap Y = \{z : Xz\ \&\ Yz\}$ and $\{x\} = \{z : z = x\}$ and $X \times Y = \{(x, y) : Xx\ \&\ Yy\}$ and $\emptyset = \{x : x \neq x\}$. In general, we use $\{x : \Phi(x)\}$ as an abbreviation for the concept F such that Fx iff $\Phi(x)$, assuming that $\Phi(x)$ is a formula which falls under one of the comprehension principles available in the theory in which we are working.

This weak background theory Σ_1^1-$\mathtt{OS}$ of second-order logic is used to define the following Fregean theory at issue in this paper. If $E(R, S)$ is an L_0-formula with two free n_E-ary concept variables and no further free variables, then we let $\mathrm{Equiv}(E)$ abbreviate the L_0-sentence expressive of E being an equivalence relation on n_E-ary concepts, i.e. the universal closure of the following, wherein R, S, T are n_E-ary concept variables:

(2.7) $$[E(R, R)\ \&\ (E(R, S) \rightarrow E(S, R))\ \&\ ((E(R, S)\ \&\ E(S, T)) \rightarrow E(R, T))]$$

Then consider the following collection of L_0-formulas which consists of all the L_0-formulas $E(R, S)$ with two free n_E-ary concept variables and no further free variables such that Σ_1^1-$\mathtt{OS}$ proves $\mathrm{Equiv}(E)$:

(2.8) $$\mathrm{ProvEquiv}(L_0) = \{E(R, S) \text{ is an } L_0 \text{ formula} : \Sigma_1^1\text{-}\mathtt{OS} \vdash \mathrm{Equiv}(E)\}$$

Then define the following expansion of L_1 of L_0:

Definition 2.7. *Let L_1 consist of the expansion of the signature L_0 (2.1) by a new function symbol ∂_E from n_E-ary concepts to objects for each E from $\mathrm{ProvEquiv}(L_0)$ (2.8).*

Then we define the predicative theory as follows:

Definition 2.8. *The* predicative Fregean theory, *abbreviated* PFT, *is the L_1-theory consisting of (i) the extensionality axioms (2.5) and the projection axioms (2.6) and (ii) the Σ_1^1-Choice Schema for L_1-formulas (Definition 2.5) and (iii) the First-Order Comprehension Schema for L_1-formulas (Definition 2.3), and (iv) the abstraction principle $A[E]$ (2.1) for each E from $\mathrm{ProvEquiv}(L_0)$ (2.8).*

Hence, the theory PFT is a recursively enumerable theory in a recursively enumerable signature L_1. If one desired a recursive signature, one could alternatively define L_1 to consist of function symbols ∂_E from n_E-ary concepts to objects for *each* L_0-formula E, regardless of whether it was in $\mathrm{ProvEquiv}(L_0)$ (2.8). This is because clause (iv) in Definition 2.8 only includes the abstraction principle $A[E]$ (2.1) when the formula E is in fact in the set $\mathrm{ProvEquiv}(L_0)$ (2.8).

While this definition is technically precise, the niceties ought not obscure the intuitiveness of the motivating idea. For, the idea behind this predicative Fregean theory is that it conjoins traditional predicative constraints on comprehension together with the idea that abstraction principles associated to certain L_0-formulae are always available. More capaciously: if we start from weak background theory of second-order logic Σ_1^1-OS and if we can prove in this theory that an L_0-formula $E(R,S)$ in the signature of this weak background logic is an equivalence relation on n_E-ary concepts for some $n_E \geq 1$, then the predicative Fregean theory PFT includes the abstraction principle $A[E]$ (2.1) associated to E. Hence the theory PFT includes the abstraction principles associated to number, extension, and ordinal, namely Hume's Principle (1.1), Basic Law V (1.4) and the abstraction principle associated to ordinals (cf. (2.2) above).

One of the aims of the earlier paper [31] was to establish the following:

Theorem 2.9. *The theory* PFT *is consistent.*

Proof. Let $E_1, \ldots, E_n, \ldots$ enumerate the elements of the collection $\mathrm{ProvEquiv}(L_0)$ from equation (2.8). By compactness, it suffices to establish, for each $n \geq 1$, the consistency of the subsystem of PFT which is formed by restricting part (iv) of the Definition of PFT to the abstraction principles $A[E_1], \ldots, A[E_n]$. But then this theory is a subtheory of the theory which, in the paper [31], we called $\Sigma_1^1 - [\mathtt{E_1}, \ldots, \mathtt{E_n}]\mathtt{A} + \mathtt{SO} + \mathtt{GC}$. The consistency of this theory was established in the Joint Consistency Theorem of that paper. $\square$

3 Interpreting second-order arithmetic in the theory

While the predicative Fregean Theory only explicitly includes predicative instances of the comprehension schema for L_0-formulas, surprisingly it is able to deductively recover all instances of the Full Comprehension Schema for L_0-formulas.

Theorem 3.1. PFT *proves each instance of the Full Comprehension Schema for L_0-formulas.*

Proof. Let $\Phi(x, G)$ be an L_0-formula with all free variables displayed, wherein x is an object variable and G is a unary concept variable. Let us first show that `PFT` proves the following instance of the Full Comprehension Schema for L_0-formulas (Definition 2.2):

$$(3.1) \qquad \forall\, G\, \exists\, F\, \forall\, x\, (Fx \leftrightarrow \Phi(x, G))$$

After we finish the proof of this instance, we'll comment on how to establish the general case.

First consider the following L_0-formulas $\mu(R, S), \nu(R, S)$ with all free variables displayed, where R, S are binary concept variables:

$$\mu(R, S) \equiv [\exists\,!\, x, G \text{ with } R = \{x\} \times G]\ \&\ [\exists\,!\, y, H \text{ with } S = \{y\} \times H]$$
$$\&\ \forall\, x, G, y, H\ [(R = \{x\} \times G\ \&\ S = \{y\} \times H) \rightarrow (\Phi(x, G) \leftrightarrow \Phi(y, H))]$$
$$\nu(R, S) \equiv \neg[\exists\,!\, x, G \text{ with } R = \{x\} \times G]\ \&\ \neg[\exists\,!\, y, H \text{ with } S = \{y\} \times H]$$

In this, the identity $R = \{x\} \times G$ is an abbreviation for the claim that

$$(3.2) \qquad \forall\, a, b\ (R(a, b) \leftrightarrow ((a = x)\ \&\ Gb))$$

Hence, $\mu(R, S)$ expresses that R can be written uniquely as $\{x\} \times G$ for some x, G, while S can be written uniquely as $\{y\} \times H$ for some y, H, and that $\Phi(x, G) \leftrightarrow \Phi(y, H)$. The circumstance in which a binary relation R can be written as $\{x\} \times G$ but not *uniquely* so is when G is empty, since in this case $\{x\} \times G = \{x'\} \times G$ for any objects x, x'. Finally, consider the following L_0-formula $E(R, S)$ where again R, S are binary concept variables and all free variables are displayed:

$$(3.3) \qquad E(R, S) \equiv (\mu(R, S) \vee \nu(R, S))$$

The weak background theory $\Sigma_1^1\text{-OS}$ proves that $E(R, S)$ is an equivalence relation on binary concepts. For reflexivity, either R can be written uniquely as $\{x\} \times G$ for some x, G, or not. If so, then one trivially has $\Phi(x, G) \leftrightarrow \Phi(x, G)$. This then implies $\mu(R, R)$ and so $E(R, R)$. If not, then of course $\nu(R, R)$ and so $E(R, R)$. For symmetry, it simply suffices to note that both μ and ν are symmetric in that $\mu(R, S)$ implies $\mu(S, R)$ and likewise for ν. For transitivity, suppose that $E(R, S)$ and $E(S, T)$. Because of the disjunctive definition of E in (3.3), there are three cases to consider. First suppose that $\mu(R, S)$ and $\mu(S, T)$. Then we may uniquely write $R = \{x\} \times G, S = \{y\} \times H, T = \{z\} \times I$, and from $\Phi(x, G) \leftrightarrow \Phi(y, H)$ and $\Phi(y, H) \leftrightarrow \Phi(z, I)$ we may conclude that $\Phi(x, G) \leftrightarrow \Phi(z, I)$. Hence we then have $\mu(R, T)$ and thus $E(R, T)$. Second suppose that $\nu(R, S)$ and $\nu(S, T)$. These two assumptions imply that we can't write any of R, S, T uniquely as the product of a singleton and a unary concept, and hence that $\nu(R, T)$ and $E(R, T)$. Finally, suppose that $\mu(R, S)$ and $\nu(S, T)$ (or vice-versa). But this case leads to a contradiction, since $\mu(R, S)$ implies that we can write S uniquely as the product of a singleton and a unary concept, while $\nu(S, T)$ says that we can't. Hence $E(R, S)$ is indeed an equivalence relation on binary concepts, and provably so in the weak background theory $\Sigma_1^1\text{-OS}$.

Then the L_0-formula $E(R, S)$ is in the set $\mathrm{ProvEquiv}(L_0)$ (2.8). Hence the theory $\mathtt{PFT}$ contains the abstraction principle $A[E]$ (2.1). Before we verify (3.1), let us introduce another abstraction principle. Consider the following L_0-formulas $\mu'(X, Y), \nu'(X, Y)$ with all free variables displayed, where X, Y are unary concept variables:

$$\mu'(X, Y) \equiv \exists\, x\, \exists\, y\; X = \{x\}\; \&\; Y = \{y\}\; \&\; (\Phi(x, \emptyset) \leftrightarrow \Phi(y, \emptyset))$$
$$\nu'(X, Y) \equiv \neg(\exists\, x\; X = \{x\})\; \&\; \neg(\exists\, y\; Y = \{y\})$$

Then consider the following L_0-formula $E'(X, Y)$ where again X, Y are unary concept variables and all free variables are displayed:

(3.4) $$E'(X, Y) \equiv (\mu'(X, Y) \vee \nu'(X, Y))$$

By the same argument as the previous paragraph, $\Sigma_1^1\text{-}\mathtt{OS}$ proves that $E'(X, Y)$ is an equivalence relation unary concepts. So the theory $\mathtt{PFT}$ contains the abstraction principle $A[E']$ (2.1)

Now, working in $\mathtt{PFT}$, let us verify (3.1). There are three cases. First suppose that there is no x_0 with $\Phi(x_0, G)$. Then to establish (3.1) one can take $F = \emptyset$.

As a second case, suppose that there is a x_0 with $\Phi(x_0, G)$ and that G is non-empty. Then observe that the graph of the function $f(x) = \partial_E(\{x\} \times G)$ has both a Σ_1^1- and a Π_1^1-definition:

$$f(x) = y \leftrightarrow \exists\, R\; (\forall\, a, b\; R(a, b) \leftrightarrow (a = x\; \&\; Gb))\; \&\; \partial_E(R) = y$$
(3.5) $$\leftrightarrow \forall\, R\; (\forall\, a, b\; R(a, b) \leftrightarrow (a = x\; \&\; Gb)) \rightarrow \partial_E(R) = y$$

These are equivalent because we can use the First-Order Comprehension Schema for L_1-formulas to secure that the binary relation $R = \{x\} \times G$ exists. Hence by the Δ_1^1-Comprehension Schema for L_1-formulas, the equivalence in (3.5) implies that the graph of f exists as a binary concept. Then by First-Order Comprehension Schema for L_1-formulas, the following unary concept exists:

(3.6) $$F = \{x : f(x) = \partial_E(\{x_0\} \times G)\}$$

Now let's argue that $F = \{x : \Phi(x, G)\}$. First suppose that Fx. Then $f(x) = \partial_E(\{x_0\} \times G)$ and hence $\partial_E(\{x\} \times G) = \partial_E(\{x_0\} \times G)$. Then $E(\{x\} \times G, \{x_0\} \times G)$ and since G is non-empty we have $\mu(\{x\} \times G, \{x_0\} \times G)$. Then $\Phi(x, G) \leftrightarrow \Phi(x_0, G)$. Since we're assuming that $\Phi(x_0, G)$, we then conclude that $\Phi(x, G)$, which is what we wanted to show. For the converse, suppose that $\Phi(x, G)$. Since we're assuming that $\Phi(x_0, G)$ and that G is non-empty we may conclude that $\mu(\{x\} \times G, \{x_0\} \times G)$ and thus $E(\{x\} \times G, \{x_0\} \times G)$ and $\partial_E(\{x\} \times G) = \partial_E(\{x_0\} \times G)$. By the definition of f, we then have $f(x) = \partial_E(\{x_0\} \times G)$ which by the definition of F implies that Fx, which is what we wanted to show.

As a third case, suppose that there is an x_0 with $\Phi(x_0, G)$ but that G itself is empty. Then we argue as before that the graph of $g(x) = \partial_{E'}(\{x\})$ exists as a binary concept, that $F = \{x : g(x) = \partial_{E'}(\{x_0\})\}$ exists as a unary concept, and that $F = \{x : \Phi(x, G)\}$.

This finishes the proof of (3.1) in $\mathtt{PFT}$. The proof of the general case of the Full Comprehension Schema for L_0-formulas (Definition 2.2) differs only in that unary concept variable F

10

from (3.1) might instead be an n-ary concept variable and there may be more than one concept parameter G, as well as some additional object parameters. But the proof of this general case is directly analogous to the proof of (3.1). The only difference is that the number of abstraction principles used in the proof will increase with the number of concept parameters. In general if there are m-concept parameters $G_1, \ldots, G_m$, then there will be 2^m different abstraction principles used in the proof, since one must consider a case corresponding to the finite binary sequence $(i_1, \ldots, i_m)$, wherein $i_k = 0$ indicates that G_k is empty, and $i_k = 1$ indicates that G_k is non-empty. $\qquad\square$

Before turning to the proof that PFT interprets second-order Peano arithmetic, let's briefly note that in the consistency proof from [31] invoked in the proof of Theorem 2.9, we explicitly verified the Full Comprehension Schema for L_0-formulas. (In the language of that paper, these were part of the theory SO, and the interested reader may consult the proof of the Joint Consistency Theorem in that paper).

While the theory PFT only explicitly includes some instances of the Full Comprehension Schema for L_0-formulas in its definition (cf. Definition 2.8), the previous theorem says that it proves all of them. However, even in this predicative setting, the Russell paradox can be used to show that there is no concept consisting of the extensions, i.e. the range of the extension operator ∂ from Basic Law V (1.4). For a proof, see [29] Proposition 29 p. 1692. Now the formula $\text{rng}(\partial)$ is definable by a Σ_1^1-formula of the signature $L_0[\partial]$. Further $L_0[\partial]$ is included in the signature L_1 of PFT. Hence, since the L_1-theory PFT is consistent by Theorem 2.9, it follows that PFT does not prove all instances of the Full Comprehension Schema for L_1-formulas.

This kind of situation is of course not entirely unfamiliar. For instance, Presburger arithmetic yields a complete axiomatization of the structure $(\mathbb{Z}, 0, 1+, <)$ (cf. Marker [24] pp. 82 ff). So this axiomatization proves each instance of the following induction schema in the signature $L = \{0, 1, +, <\}$:

$$(3.7) \qquad [\varphi(0) \ \& \ \forall \ x \geq 0 \ (\varphi(x) \to \varphi(x+1)))] \to [\forall \ x \geq 0 \ \varphi(0)]$$

Consider a non-standard model $G = (G, 0, 1, +, <)$ of Presburger arithmetic, and extend L to L' by adding a new unary predicate Z which is interpreted on G as the integers $\mathbb{Z}$. Then of course the axioms of Presburger arithmetic do not imply all instances of the schema (3.7) in the expanded signature L'. So of course it's consistent for there to be a schema and an L'-theory and a subsignature L of L' such that the theory proves all instances of the L-schema but not every instance of the L'-schema.

Now let's show that PFT interprets second-order Peano arithmetic PA^2. These axioms are the natural set of axioms used to describe the standard model of second-order arithmetic; see [26] p. 4 or [29] p. 1680 or [30] p. 106 for an explicit list of these axioms.

Theorem 3.2. *The predicative Fregean theory* PFT *interprets second-order Peano arithmetic* PA^2.

Proof. First note that the predicative Fregean theory PFT proves the existence of the graph of the function $s(x) = \partial(\{x\})$ (cf. [29] Proposition 27 p. 1691), where this is the abstraction

11

operator associated to Basic Law V (1.4). For, note that in PFT, for all objects x, y, one has that the following Σ_1^1-condition and Π_1^1-conditions are equivalent:

$$(3.8) \qquad [\exists\ X\ (X = \{x\}\ \&\ \partial(X) = y)] \leftrightarrow [\forall\ X\ (X = \{x\} \rightarrow \partial X = y)]$$

By the Δ_1^1-Comprehension Schema for L_1-formulas, there is then a binary relation which holds of objects x, y iff either the Σ_1^1-condition holds or the Π_1^1-condition holds. And this binary relation is obviously the graph of the function $s(x) = \partial(\{x\})$.

Let M be $\{x : x = x\}$, which exists by Full Comprehension for L_0-formulas, and let $0 = \partial(\emptyset)$. Then one has that the triple $(M, 0, s)$ satisfies the first two axioms of Robinson's Q:

$$(3.9) \qquad \forall\ x\ s(x) \neq 0, \qquad \forall\ x, y\ (s(x) = s(y) \rightarrow x = y)$$

For, suppose that $s(x) = 0$. Then $\partial(\{x\}) = \partial(\emptyset)$ and then by Basic Law V (1.4) one has that $\{x\} = \emptyset$, a contradiction. Similarly, suppose that $s(x) = s(y)$. Then $\partial(\{x\}) = \partial(\{y\})$ and so by Basic Law V (1.4) one has that $\{x\} = \{y\}$ and hence $x = y$. Thus (3.9) follows immediately from Basic Law V (1.4).

But then standard arguments allow one to interpret second-order Peano arithmetic PA^2 by taking the natural numbers N to be the sub-concept of M consisting of all those subconcepts of M which are "inductive," that is which contain zero and closed under successor. Here of course for the existence of N and the verification of the other axioms of arithmetic, one appeals to the Full Comprehension Schema for L_0-formulas, using $M, 0, s$ as parameters (cf. [29] Theorem 16 p. 1688). $\qquad\square$

4   The fragility of abstraction with predicative comprehension

However, in spite of its technical strength, the conceptual basis of the predicative Fregean theory PFT is rather fragile. For, the L_1-theory PFT was formed by adding the abstraction principle $A[E]$ associated to the L_0-formulas $E(R, S)$ when this formula could be proven to be an equivalence relation in the background second-order logic Σ_1^1-OS. But one cannot successively iterate this idea. For, suppose that in analogue to ProvEquiv(L_0) in equation (2.8), one defines:

$$(4.1) \qquad \text{ProvEquiv}(L_1) = \{E(R, S) \text{ is an } L_1 \text{ formula} : \text{PFT} \vdash \text{Equiv}(E)\}$$

And further suppose that one defines L_2 to be the expansion of L_1 by the addition of a function symbol ∂_E from n_E-ary concepts to objects for each L_1-formula $E(R, S)$ in ProvEquiv(L_1). Finally, suppose one defines the following iteration of PFT (cf. Definition 2.8):

Definition 4.1. *The theory* PFT$_2$ *is the L_2-theory consisting of (i) the extensionality axioms (2.5) and the projection axioms (2.6) and (ii) the Σ_1^1-Choice Schema for L_2-formulas (Definition 2.5) and (iii) the First-Order Comprehension Schema for L_2-formulas (Definition 2.3),*

and (iv) the abstraction principle $A[E]$ (2.1) for each E which is from $\mathrm{ProvEquiv}(L_0)$ (2.8) or from $\mathrm{ProvEquiv}(L_1)$ (4.1).

Then the same argument as in the proof of Theorem 3.1 establishes that $\mathtt{PFT}_2$ proves each instance of the Full Comprehension Schema for L_1-formulas. But then $\mathtt{PFT}_2$ is inconsistent, since on pain of the Russell paradox there is no concept of all extensions (cf. [29] Proposition 29 p. 1692), where again the extensions are the range of the abstraction operator ∂ associated to Basic Law V (1.4). Hence, while the predicative Fregean theory $\mathtt{PFT}$ is consistent, when one tries to iterate its underlying idea of adding abstraction principles when their equivalence relations can be proven to be equivalence relations, one again runs up against the Russell paradox. This indicates that the resource of abstraction principles in the predicative setting is unlike that of typed theories of truth or second-order logic, which we may consistently add to any consistent theory.

This point is underscored when one observes that the same considerations show the inconsistency of an axiom-based analogue of the rule-based predicative Fregean theory $\mathtt{PFT}$. In particular, suppose that we recursively defined a signature L^* extending L_0 so that if $E(R, S)$ is an L^*-formula in exactly two free n_E-ary concept variables then L^* also contains a function symbol ∂_E which takes n_E-ary concepts to objects and which does not occur in E. One could then define the following L^*-theory:

Definition 4.2. *The theory $\mathtt{PFT}^*$ is the L^*-theory consisting of (i) the extensionality axioms (2.5) and the projection axioms (2.6) and (ii) the Σ^1_1-Choice Schema for L^*-formulas (Definition 2.5) and (iii) the First-Order Comprehension Schema for L^*-formulas (Definition 2.3), and (iv) the axiom $\mathrm{Equiv}(E) \to A[E]$ for each L^*-formula E.*

In this, $\mathrm{Equiv}(E)$ is the sentence which says that E is an equivalence relation (cf. (2.7)) and $A[E]$ is the abstraction principle (2.1), so that the axiom $\mathrm{Equiv}(E) \to A[E]$ says that if E is an equivalence relation, then $A[E]$ holds. The considerations of the previous paragraphs can be replicated in this theory $\mathtt{PFT}^*$, showing it to be inconsistent. However, the conceptual distance between the inconsistent L^*-theory $\mathtt{PFT}^*$ and the consistent L_1-theory $\mathtt{PFT}$ is rather slim. The difference is merely a difference between a rule and an axiom: whereas the rule-based $\mathtt{PFT}$ only includes an abstraction principle when the underlying equivalence relation is expressible in the weak background logic and is *provably* an equivalence relation there, the axiom-based $\mathtt{PFT}^*$ includes a commitment to either the truth of the abstraction principle or the falsity of its underlying formula being an equivalence relation.

In response to this, one might try to restrain the predicative Fregean theory $\mathtt{PFT}$ so that the analogously defined iterated version of it and the analogously defined axiom-based version of it were consistent. For instance, one might consider restricting the abstraction principles added to the theory $\mathtt{PFT}$ to those whose underlying equivalence relation was expressible both as a Σ^1_1-formula and a Π^1_1-formula in the background second-order logic. This, it might be suggested, would be a genuinely predicative theory of abstraction principles. Such a move would block the proof of Theorem 3.1. For, the equivalence relation $E(R, S)$ (3.3) used in that proof is not obviously expressible in such a way. However, it is unknown to us how much arithmetic this more austerely predicative theory could interpret, and it is not obvious

to us whether the analogously defined iterated version of it (or axiom-based version of it) is consistent.

Another way forward might be to find some principled way to focus attention on abstraction principles which are somehow more like the paradigmatic Basic Law V (1.4) and Hume's Principle (1.1) and the abstraction principle associated to ordinals (2.2), and somehow less like the seemingly ad-hoc abstraction principles constructed in the proof of Theorem 3.1. But to do so would be to lose some of the original motivation for focusing on predicative abstraction principles. For, part of the attraction was supposed to be that more abstraction principles became consistent and jointly consistent. And indeed, as the predicative Fregean Theory PFT attests, a good deal of joint consistency is available in this setting. Hence in the earlier paper [31] we said that we had resolved an analogue of the joint consistency problem. But as we have seen in this section, when we try to iterate the underlying idea of abstraction principles in the predicative setting, we again run into inconsistency and seem back in the situation of trying to discern ways to weed out the acceptable from the unacceptable abstraction principles. For an overview of the various candidates for acceptable abstraction principles in the general impredicative setting, see [23] or [5].

Perhaps another way forward might be to give up on the idea of abstraction principles altogether and find principled reasons for studying systems centered around either Basic Law V (1.4) itself or Hume's Principle (1.1) itself or the abstraction principle associated to ordinals (2.2) all by itself. With respect to Basic Law V (1.4), this is the perspective of [32], where the idea is to work within an intensional logic and see the extension operator as selecting a sense for each concept, just like we might select a specific Turing machine index for each computable function. But much remains unknown about the individual abstraction principles at the predicative level. For instance, it is to our knowledge unknown whether Basic Law V (1.4) or the abstraction principle associated to ordinals (2.2), equipped with the Σ_1^1-choice schema and the First-Order Comprehension Schema, interprets the analogous predicative versions of arithmetic (cf. [29] p. 1707). In this paper, the idea for interpreting arithmetic was to collect together all the predicative abstraction principles so that they could effect the interpretation together, and it is in general unclear to us what happens when one focuses on the abstraction principles one by one.

Acknowledgements

References

[1] Evert W. Beth. Chapter 13: Logicism. In *The Foundations of Mathematics: A Study in the Philosophy of Science*, Studies in Logic and the Foundations of Mathematics, pages 353–364. North-Holland, Amsterdam, 1959. 2

[2] George Boolos. *Logic, logic, and logic*. Harvard University Press, Cambridge, MA, 1998. Edited by Richard Jeffrey. 5

[3] John P. Burgess. *Fixing Frege*. Princeton Monographs in Philosophy. Princeton University Press, Princeton, 2005. 3, 7

[4] Roy T. Cook, editor. *The Arché papers on the mathematics of abstraction*, volume 71 of *The Western Ontario Series in Philosophy of Science*. Springer, Berlin, 2007. 2

[5] Roy T. Cook. Conservativeness, stability, and abstraction. *British Journal for the Philosophy of Science*, 63:673–696, 2012. 14

[6] William Demopoulos, editor. *Frege's philosophy of mathematics*. Harvard University Press, Cambridge, 1995. 16

[7] Solomon Feferman. Systems of predicative analysis. *The Journal of Symbolic Logic*, 29:1–30, 1964. 2

[8] Solomon Feferman. Predicativity. In Stewart Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–624. Oxford University Press, Oxford, 2005. 2

[9] Fernando Ferreira and Kai F. Wehmeier. On the consistency of the Δ^1_1-CA fragment of Frege's Grundgesetze. *Journal of Philosophical Logic*, 31(4):301–311, 2002. 3

[10] Gottlob Frege. *Die Grundlagen der Arithmetik*. Koebner, Breslau, 1884. 2

[11] Gottlob Frege. *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*. Pohle, Jena, 1893, 1903. Two volumes. Reprinted in [12]. 2

[12] Gottlob Frege. *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*. Olms, Hildesheim, 1962. 15

[13] Gottlob Frege. *The foundations of arithmetic: a logico-mathematical enquiry into the concept of number*. Northwestern University Press, Evanston, second edition, 1980. Translated by John Langshaw Austin. 2

[14] Gottlob Frege. *Basic laws of arithmetic*. Oxford University Press, Oxford, 2013. Translated by Philip A. Ebert and Marcus Rossberg. 2

[15] Harvey M. Friedman. Some systems of second-order arithmetic and their use. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974*, volume 1, pages 235–242. 1975. 2

[16] Mihai Ganea. Burgess' *PV* is Robinson's *Q*. *Journal of Symbolic Logic*, 72(2):618–624, 2007. 3

[17] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Perspectives in Mathematical Logic. Springer, Berlin, 1998. 2, 3

[18] Bob Hale and Crispin Wright. *The reason's proper study*. Oxford University Press, Oxford, 2001. 2

[19] Richard G. Heck, Jr. The consistency of predicative fragments of Frege's Grundgesetze der Arithmetik. *History and Philosophy of Logic*, 17(4):209–220, 1996. 3

[20] Richard G. Heck, Jr. *Frege's theorem*. Oxford University Press, Oxford, 2011. 2

[21] Harold Hodes. Logicism and the Ontological Commitments of Arithmetic. *The Journal of Philosophy*, 81(3):123–149, 1984. 5

[22] Per Lindström. *Aspects of incompleteness*, volume 10 of *Lecture Notes in Logic*. Association for Symbolic Logic, Urbana, IL, second edition, 2003. 2

[23] Øystein Linnebo. Some criteria for acceptable abstraction. *Notre Dame Journal of Formal Logic*, 52(3):331–338, 2010. 14

[24] David Marker. *Model Theory: An Introduction*, volume 217 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. 11

[25] Terence Parsons. On the consistency of the first-order portion of Frege's logical system. *Notre Dame Journal of Formal Logic*, 28(1):161–168, 1987. Reprinted in [6]. 3

[26] Stephen G. Simpson. *Subsystems of second order arithmetic*. Cambridge University Press, Cambridge, second edition, 2009. 2, 3, 5, 6, 7, 11

[27] Albert Visser. Categories of theories and interpretations. In Ali Enayat, Iraj Kalantari, and Mojtaba Moniri, editors, *Logic in Tehran*, volume 26 of *Lecture Notes in Logic*, pages 284–341. Association for Symbolic Logic, La Jolla, 2006. 2

[28] Albert Visser. The predicative Frege hierarchy. *Annals of Pure and Applied Logic*, 160(2):129–153, 2009. 3

[29] Sean Walsh. Comparing Hume's principle, Basic Law V and Peano Arithmetic. *Annals of Pure and Applied Logic*, 163:1679–1709, 2012. 2, 3, 4, 5, 6, 11, 12, 13, 14

[30] Sean Walsh. Logicism, interpretability, and knowledge of arithmetic. *The Review of Symbolic Logic*, 7(1):84–119, 2014. 2, 3, 11

[31] Sean Walsh. Fragments of Frege's *Grundgesetze* and Gödel's constructible universe. *The Journal of Symbolic Logic*, forthcoming. http://arxiv.org/abs/1407.3861. 3, 4, 8, 11, 14

[32] Sean Walsh. Predicativity, the Russell-Myhill paradox, and Church's intensional logic. *The Journal of Philosophical Logic*, forthcoming. http://dx.doi.org/10.1007/s10992-015-9375-5, http://arxiv.org/abs/1506.02206. 3, 4, 14

[33] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis.* Veit, Leipzig, 1918. 2

[34] Crispin Wright. *Frege's conception of numbers as objects*, volume 2 of *Scots Philosophical Monographs*. Aberdeen University Press, Aberdeen, 1983. 2